Pitching an Argument: Intonation, information, and inference in syllogistic discourse

Stuart Ian Hughson

Doctor of Philosophy
Institute for Communicating and Collaborative Systems
School of Informatics
University of Edinburgh
2007

Abstract

In the century or so that syllogisms have received the attention of psychologists, their interpretation both as and within particular types of discourse has been largely downplayed. A significant element in the guidance of interpretation is information structure as expressed in prosody. This thesis explores the rôle of intonation in syllogistic discourse and its effect on reasoning.

A theoretical analysis of the information structures of syllogisms is presented which produces two classes of intonation patterns, the 'contextually concordant' (*CC*) and the 'contextually neutral' (*CN*), putatively corresponding to two discourse types. These are then investigated in a series of experiments. The initial observational study aims at confirming the use and significance of *CC* and *CN* patterns in a syllogism solving task. The remaining two experiments employ a purpose-built voice synthesiser to investigate the effects of imposing *CC* and *CN* contours on premises, first in a syllogism solving task and then in a syllogism evaluation task.

The results show that both *CC* and *CN* intonation patterns are indeed used by participants and bear a systematic relationship to both the number and accuracy of conclusions they draw. When used in the presentation of syllogisms, however, these patterns do not influence the production of conclusions, only the evaluation of them. It is therefore argued that the discourse types to which they relate depend upon whether the syllogism is interpreted as a proof or as a problem.

Further work based on these findings could aim to probe the informational links between conclusions and premises and thereby elucidate the coherence of arguments.

Acknowledgements

The twin pillars that have kept this thesis from crashing to the ground are my two supervisors, Professors Jon Oberlander and Keith Stenning. Aside from his Mycroftian range, Jon has an unfailing knack for delivering exactly the right combination of supportive encouragement and a kick in the pants. I've needed plenty of both. Keith's genius is for confusing the hell out of me in a way I can't simply let lie. Much of my intellectual life in Edinburgh has been shaped by the academic koans he has given me. I am particularly thankful to both of them for their forbearance during what proved to be a very troublesome writing up phase.

I have also been extremely fortunate in benefiting from our close links with the School of Philosophy, Psychology, and Language Sciences. Professor Bob Ladd gave generously of both his wisdom and expertise and his department's resources. The Festival team answered my e-mails even over Christmas. I have to single out Rob Clark, though, for his unfailingly patient and informatively detailed advice in the face of my constant pestering over months of trying to get the speech synthesiser to do what I wanted.

Obviously, this thesis would have been nothing without the experiments, but the experiments would have been nothing without Alan Whyte, Eddie Dubourg, Ziggy, and Mike Bennett, all of whom seem permanently willing to do more than is asked of them. I'd also like to thank my independent annotators, Amy Isard and Alistair Gill, for being willing to work for nothing other than proving I still have some marbles left. The credit for actually preserving them, though, along with my eternal gratitude, goes to Frank Keller for pointing me in the direction of |STAT.

It is the perverse lot of those dearest to us that they come last in lists such as this. Perhaps it is tacit acknowledgement of the place we short-sightedly consign them to as our labours come to dominate our lives. Maria had the least and the worst of me throughout, for which I am truly sorry.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Stuart Ian Hughson*)

Table of Contents

Chapter 1

Introduction

It is almost a century since psychologists first turned their attention to the ancient form of argument known as the syllogism.[1] Experimental participants with no training in logic, while performing better than chance, make many mistakes in syllogistic reasoning tasks, according to the definitions of classical logic. Over the years, a variety of extra-logical factors has been found to influence their performance. (See Chapter 2 for an overview). For the most part, these results have been taken as evidence for or against different theories of the nature of the human reasoning mechanism. Perhaps it is the pared down, systematised characterisation of them that encourages this, but in concentrating thus on syllogisms as paradigms of inference these studies have largely ignored them as models of discourse. As a result, two significant issues have been taken for granted and therefore sidelined, namely (1) that the purpose of a syllogism is to demonstrate deductively sound reasoning and (2) that experimental participants understand and aim for this. There are, however, very good reasons for believing that neither of these is necessarily true.

Stenning and van Lambalgen (2005) distinguish between 'skeptical' and 'credulous' attitudes which lead to very different interpretations of and within discourse. The former is akin to the logical quest for truth under all possible interpretations of a given set of propositions, but with a localised, contextualised conception of the possibilities more in tune with the multiplicity of contemporary, specialised logics than with the old-fashioned, monolithic approach. The latter seeks truth only in the guise of the one, particular interpretation which it is the assumed intention of a discourse partner to convey. In the latter case, but not the former, there is a presumption that the statements presented are related to each other, thereby constraining the search for the correct in-

[1] Störring (1908) is believed to be the earliest study (Politzer, 2004).

terpretation. Ordinary language users switch with complete facility between these two modes of operation as occasion demands: a credulous attitude serves for the reception of information for so long as no doubt or confusion as to the intended model arises, at which point a skeptical view that suspends earlier assumptions and contemplates alternative models assists the process of repair.

This conception of the difference between exposition and deduction as types of discourse has formed the basis of accounts of reasoning with conditionals in respect of both the suppression of *modus ponens* (Byrne, 1989; Stenning and van Lambalgen, 2005) and variation in performance of Wason's selection task (Wason, 1966; Stenning and van Lambalgen, 2004). In the latter case, the claim is that the task creates a situation calling for skeptical reasoning, but differences in the content of the materials evoke logics tailored to different needs. The problem confronting the participant is usually compounded, however, by an inadequate specification, if not understanding, of the nature of the task by the experimenter. They are therefore at liberty to form a conception of the discourse in question that is at odds with that held by the experimenter, with the result that the responses given by the one are misconstrued by the other.

Clearly, the potential for such natural reinterpretation of the experimental discourse extends beyond these two paradigms. Indeed, there is plenty of evidence that studies of syllogistic reasoning have been subject to very much the same difficulties. Participants have long been known to perform better with materials cast in real-world terms than with abstract ones (Wilkins, 1928), but Wason and Johnson-Laird (1972) go so far as to suggest that this is in fact due to the contextualisation of inference by the problem statement itself. The well-established tendency for participants to generate or accept as valid conclusions they believe to be true and reject those they believe to be false (Janis and Frick, 1943; Morgan and Morton, 1944; Evans *et al.*, 1983; Oakhill and Johnson-Laird, 1985; Oakhill *et al.*, 1989) also fits this picture. Indeed, Henle (1962) found that syllogisms embedded in everyday contexts not only led participants to fail to distinguish between validity and truth, but also to restate premises according to their own judgements and even to reject the task itself as defined to them.

More specifically, though, Stenning and Cox (2006) suggest that, in the context of syllogisms, exposition and deduction are reflected in 'rash' and 'hesitant' inference patterns which are mediated by sensitivity to 'information packaging' (Vallduví, 1992). It is this, they argue, that variously triggers or inhibits the drawing of implicatures (Grice, 1975, 1989), leading to a complex set of interactions that influence the form and, to a limited extent, accuracy of conclusions drawn by participants. Their

account is constrained, however, by a view of information packaging restricted to subject-predicate ordering, or 'grammar', rather than more orthodox information structural constructs such as *topic/comment* and *focus/ground*, which have a discourse-level aspect. The claim of this thesis is that such constructs offer a more integrated account of the difference between the discourse types involved which addresses more fully the degree and accuracy of conclusion-drawing by participants. This claim is pursued in the remaining chapters as follows.

Chapter 2 is an introduction to syllogisms, first from a historico-logical perspective and then from a psychological one. It begins with a grounding in the forms and properties of syllogisms, along with the terminology used to refer to and describe them in the rest of the thesis. This is followed by an overview of psychological studies involving syllogistic reasoning tasks and the major effects that have been identified therein. The final section considers in greater detail Stenning and Cox's (2006) explanation of some of these effects in terms of credulous and skeptical interpretations of the experimental discourse and their relation to information packaging. While the general idea of a kind of 'para-Gricean' pragmatics of deduction is accepted, however, it is argued that an important aspect of information structure, namely intonation, has been neglected.

Chapter 3 attempts to elucidate the elusive concepts underlying the term 'information structure' and their manifestation in prosody and apply the resulting insights to the syllogistic form. It begins with something of a mystery tour through the terminological and definitional jungle that covers the territory. Underlying notions of constituency and context are used as tools to assist in teasing out the relationships between various theories and clarifying the terms used. This enables Stenning and van Lambalgen's distinction between expositive and deductive discourse to be recast entirely at the level of information structure by way of two putative classes of intonation patterns, the 'contextually concordant' (*CC*) and the 'contextually neutral' (*CN*). These two classes are then populated on the basis of a systematic analysis of the information structures of the complete set of syllogisms.

In Chapter 4, the hypothesis that credulity and skepticism are reflected in *CC* and *CN* patterns, respectively, is put to the test. The concepts of focus and topic are operationalised for the purpose of empirical investigation and applied in a study of the syllogism solving task in which participants' own prosodic structuring of premises is elicited. Both classes of intonation patterns are observed in use and shown to be systematically related to participants' conclusion drawing behaviour. Rather than delineating individual differences in interpretation between participants, however, they

appear to distinguish across the board between the different outcomes of reasoning, namely whether or not a conclusion is drawn and, in either case, whether or not this is logically correct.

Chapter 5 is a digression into the construction of an artificial voice for the purpose of incorporating controlled intonation into subsequent experiments. The process of building a unit selection voice for the FESTIVAL speech synthesis system is outlined and the application of each step to the domain of syllogisms is detailed. In particular, the choice of vocabulary and optimisation of the voice's coverage of both *CC* and *CN* patterns is explained. Attention is also given to problems with the system arising out of the exacting requirements of the application and the modifications necessary to their solution.

Chapter 6 reports an experiment to determine whether or not intonational structuring can be used to influence the conclusion drawing behaviour of participants. In an adapted syllogism solving task, the artificial voice was used to present premises variously with *CC* and *CN* contours. The results show no difference whatsoever in the number or accuracy of conclusions given in response to the two classes of intonation patterns. It is suggested that the difference between these results and those of the initial study might be that intonation forming part of the product of reasoning is not powerful enough to influence the process of reasoning.

In Chapter 7, a revised hypothesis is tested. A further experiment using the artificial voice is described, this one being an adaptation of the syllogism conclusion evaluation task. The results this time show a significant, if modest, effect of intonation pattern on the accuracy of participants' judgements. While giving possible support to the salience of the process/product distinction, though, it also raises questions concerning confidence in and believability of conclusions.

Chapter 8 concludes the thesis. The results of the three experiments are drawn together to support a revision of Stenning and van Lambalgen's distinction between exposition and deduction, at least in the context of syllogistic reasoning, into one between demonstration and examination. It is also suggested that much of the research in this field may in fact have less to do with logic in itself than with the complete *trivium* of logic, grammar, and rhetoric. Finally, proposals for investigations following on from the results reported here are put forward. Aside from consolidatory work, these include refinement of the concept of demonstration by exploring further the prosody of conclusions.

Chapter 2

Syllogistic reasoning: logic and psychology

2.1 A brief history of syllogisms

2.1.1 Aristotle

The syllogism as a form of argument or proof was first introduced and analysed by Aristotle in his *Prior Analytics*, building on the work on classes and predication in his *Categories* and on propositions in his *On Interpretation*. It consisted of just two premises which together related three terms by having one, the 'middle', in common, such that a conclusion relating the other two terms, or 'extremes', necessarily followed. For example: *Some birds are swans* and *All swans are white*, so *Some birds are white*.

The premises and conclusion were all simple propositions which predicated one term of another, either wholly or in part. Propositions could be universal or particular,[1] and affirmative or negative, giving rise to four types. For example:

Universal Affirmative: *All swans are white*. (Or *Every swan is white*).

Universal Negative: *No ravens are white.* (Or *No raven is white*).

Particular Affirmative: *Some birds are swans.* (Or *Some bird is a swan*).

Particular Negative: *Some birds are not swans.* (Or *Some bird is not a swan*).

[1] In fact, Aristotle contemplated a third possibility: the indefinite. This, however, was simply the case in which the proposition was not explicitly stated as universal or particular and therefore its scope was unclear. The logical consequences of such propositions always turned out to fall in line with one or other of the explicit cases and so were unproductive. Later treatments effectively ignored them.

In addition, each premise might have the middle as either subject or predicate, with the extreme being the other. This meant that the terms in the two premises taken together could appear in any one of three possible arrangements, or 'figures'. In the first figure, the middle was the subject in one premise and the predicate in the other. In the second, the middle was the predicate in both premises, while in the third, it was the subject in both. For example:

First figure: *Some birds are hawks.*
All hawks are hunters.
(Hence, *Some birds are hunters.*)

Second figure: *All swans are birds.*
No horses are birds.
(Hence, *No swans are horses.*)

Third figure: *All swans are birds.*
All swans are white.
(Hence, *Some birds are white.*)

The substance of Aristotle's examination consisted of working through every combination of figure and type of predication and showing which ones led to necessary conclusions and which did not. He considered 'perfect' those syllogisms in which the conclusion followed plainly from the premises, without further elaboration, and found that only syllogisms in the first figure satisfied this condition.[2] Syllogisms in the second and third figures were 'imperfect' and required to be transformed into their equivalents in the first figure in order to prove their conclusions. Such transformations were effected by truth-preserving, logical manipulations of one or both of the premises, the most notable being 'conversion', in which the subject and predicate terms exchanged places, as illustrated in the following examples:

Universal affirmative: *All swans are birds* converts to *Some birds are swans*.[3]

Particular affirmative: *Some birds are pets* converts to *Some pets are birds*.

[2] The semantics of predication in this context is containment or inclusion of one class by or in another and its transitivity Aristotle took as self-evident.

[3] Aristotle's prior writings make it clear that the sorts of things that terms refer to and to which his syllogistic system applies are carefully circumscribed, in accordance with his overarching empiricism, such that the existential assumption underlying this conversion is warranted. Later, mediaeval writers invoked the doctrine of *supposition* to the same effect. Modern logicians, by contrast, are content to admit the truth of assertions concerning non-existent entities, which would invalidate it. The set-theoretic solution to this is to make the explicit assumption that there are no empty sets. All well-framed psychological experiments based on the system therefore include this assumption in some guise.

Universal negative: *No horses are goats* converts to *No goats are horses*

Particular negative: *Some birds are not swans* does not convert.

It should be noted that only the particular affirmative and the universal negative converts wholly, while the universal affirmative converts but partially and the particular negative does not convert at all.

2.1.2 Scholasticism

Aristotle, it seems, was not overly concerned with the order in which the premises were stated, as this does not affect what conclusions do or do not necessarily follow. Rather, he distinguished them on the basis of whether they contained the 'major' term or the 'minor'. Unfortunately, his definition of these is a little obscure and varies between the figures. His pupil and successor, Theophrastus, followed by the Scholastics of the Middle Ages, adopted the view that the major term was that which appeared in the conclusion as the predicate and the minor term that which appeared there as the subject. Pursuing this more grammar-oriented approach into the premises led to the reconstruction of the system in four figures instead of three, since either term could be introduced in the first place as either subject or predicate. The fourth figure comprises the 'counter-figurals': a class of syllogisms excluded from Aristotle's scheme in which the major term is introduced as a subject and the minor term is introduced as a predicate and they therefore exchange grammatical rôle s in the conclusion.[4]

As an aid to this analysis, the convention was established of presenting the premises, named by reference to the terms they introduced, in the order major, then minor. In the examples that follow, they are described for illustrative purposes by surscripts in which M denotes the middle term, P the major term, S the minor term, and an arrow signifies that the term at its tail is predicated of the term at its head.

First figure: $\overset{\text{M} \longleftarrow \text{P}}{\textit{All hawks are hunters.}}$
$\overset{\text{S} \longleftarrow \text{M}}{\textit{Some birds are hawks.}}$
$\overset{\text{S} \longleftarrow \text{P}}{\text{(Hence, } \textit{Some birds are hunters.}\text{)}}$

[4] It would be wrong to think that Aristotle simply overlooked these. He considered them as part of his examination of his first figure and rejected each one for having no necessary conclusion. The precise reason for this is not clear. One possibility is that he failed to exhaust the application of his own rules of conversion to them. Perhaps more likely, though, is that his conception of the semantics of containment, particularly its asymmetry or directionality, rendered the complete inversion of terms in a conclusion nonsensical.

Second figure: *All swans are birds.* (P ⟵ M)
No horses are birds. (S ⟵ M)
(Hence, *No horses are swans.* (S ⟵ P))

Third figure: *All swans are white.* (M ⟵ P)
All swans are birds. (M ⟵ S)
(Hence, *Some birds are white.* (S ⟵ P))

Fourth figure: *Some birds are hawks.* (P ⟵ M)
All hawks are hunters. (M ⟵ S)
(Hence, *Some hunters are birds.* (S ⟵ P))

The particular sequence of proposition types in a syllogism so ordered was termed its 'mood'. With four types available to each of three propositions, there were therefore 64 potential moods. Since every mood could appear in every figure, there were in total 256 possible syllogisms. In the result, however, only 19 of these proved to be valid. These, along with their interrelationships, were memorised using sophisticated sets of mnemonics built around code letters for the four types of proposition:

A - Universal affirmative (*All*)

E - Universal negative (*No*)

I - Particular affirmative (*Some*)

O - Particular negative (*Some...not*)

This remained effectively the state of the art until term logic was abandoned in favour of mathematical logic, following the work of Boole, Frege, and Russell in the nineteenth and early twentieth centuries.

2.2 Overview of psychological investigations

2.2.1 Introduction

Psychological studies of syllogistic reasoning by untrained participants have employed several variations of the task, including deciding the validity of multiple candidate conclusions to pairs of premises (Woodworth and Sells, 1935; Sells, 1936), selecting valid conclusions from sets of alternatives (Chapman and Chapman, 1959; Dickstein,

1975), and producing conclusions given just the premises (Johnson-Laird and Steedman, 1978; Johnson-Laird and Bara, 1984). A number of them, especially the earlier ones, used only selected syllogisms and focussed on specific errors, while most of the more recent ones have presented participants with all possible variations and sought to develop more general theories of reasoning. Johnson-Laird and Steedman's (1978) study, in particular, broke new ground and broadened the scope of psychological investigations into syllogistic reasoning, by employing for the first time the task of actively drawing conclusions from pairs of premises. Since previous studies had limited themselves to the mere evaluation of syllogisms presented in their entirety according to scholastic convention, they perhaps had more to do with participants' reactions to the canons of classical logic than with their own inferential processes.

The elimination of the conclusion from the specification of a syllogism, entailed by the new task, leads to a further reformulation of the problem space. There are just 64 possible permutations of the two premises alone, but, surprisingly, 27 of them now yield logically valid conclusions. This is because the distinction between major and minor, rooted in the canonical conclusion, is lost. Participants are free to order the terms in their conclusions whichever way round they see fit and, depending upon the convertibility of the quantifier, some are valid either way while others are not. Obviously, the terms 'figure' and 'mood' do not survive this revision of the task unaltered. Figure is now best expressed as the pattern of terms in the two premises - AB-BC, BA-CB AB-CB, or BA-BC - and mood excludes the quantifier in the conclusion.

Taken together, these studies show that, on average, participants make mistakes in about half of the cases, which on the one hand is quite considerable but on the other is significantly better than chance (Dickstein, 1975; Johnson-Laird and Steedman, 1978), and that there are a number of factors that influence their performance. These factors fall into three main categories, namely those that are structural features or properties of syllogisms, those that are abstractions over the logic of syllogistic reasoning, and those that are cognitive or behavioural traits of individual participants. The most notable of these in each case are outlined in the following subsection.

2.2.2 Factors influencing syllogistic reasoning

2.2.2.1 Structural factors

To begin with, the mood of a syllogism mediates acceptance of a conclusion as valid. Certain types of premises are said to create an 'atmosphere' in which a conclusion of

the same type is more likely to be accepted than one of a different type (Woodworth and Sells, 1935; Sells, 1936). Specifically, a negative premise generates a negative atmosphere, even if the other premise is positive, and a particular premise produces a particular atmosphere, even if the other premise is universal. The effect is more pronounced for valid conclusions than for invalid ones and must therefore result from mediation of a separate inferential process (Johnson-Laird and Bara, 1984, p. 7).

Another factor affecting performance is the convertibility of premises. Participants often opt for invalid conclusions when either or both of the premises can not be fully converted. (That is, those in *All* and *Some...not*. It will be recalled that only those in *Some* and *No* convert fully.) In many cases, such erroneous responses would be correct if the premises were fully convertible (Chapman and Chapman, 1959). The explanation for this has taken a number of forms, some stronger than others, but in fact most of the evidence is in respect of *All* (see Dickstein, 1981, for a review).

Thirdly, the figure of a syllogism impacts powerfully on both the form and accuracy of conclusions drawn by participants. Premises whose terms are in the order AB-BC evoke a bias towards conclusions in the order AC while those ordered BA-CB evoke a bias towards CA conclusions.[5] That is, there is a strong tendency for terms to occupy the same positions in conclusions as they do in premises. These biases are manifest in invalid conclusions as well as valid ones. Furthermore, where a conclusion is valid in only one order, the ease with which the syllogism is solved reflects the compatibility of that order with the figure of the syllogism, both in respect of the time taken to produce a conclusion and the number of correct conclusions produced. Premises whose term ordering is symmetrical, i.e. AB-CB or BA-BC, evoke little or no bias in conclusion term ordering and are of intermediate difficulty (Johnson-Laird and Steedman, 1978).

Figure also contributes to the difficulty of a syllogism in another way. The response that there is NO VALID CONCLUSION (NVC) is elicited proportionately least from premises with terms ordered AB-BC. This is followed by BA-CB, then AB-CB, with the the order BA-BC producing the largest percentage of NVC responses. The effect is independent of the existence or non-existence of valid conclusions.[6] It occurs when participants are required to respond quickly, as well as after they have been allowed to reflect on and modify their answers, suggesting that it might be the result of failure to create a unified representation for both premises in a problem in the first place (Johnson-Laird and Bara, 1984).

[5] The latter are canonical according to scholastic logic, but the former are not.

[6] In fact, the order BA-BC generates the greatest number of logically valid conclusions.

2.2.2.2 Logical factors

The relative difficulty of different syllogisms can also be linked to an abstract property they may be considered to possess in virtue of model-based theories of reasoning, of which 'mental models' (Johnson-Laird, 1983) is the prime exemplar. According to this theory, syllogistic reasoning proceeds via three steps. First, a model of the state of affairs described by the premises is constructed. Next, an 'informative' conclusion, i.e. one that relates the end terms directly, is produced by scanning the model. Then, an alternative model is sought that remains consistent with the premises but not with the conclusion and the process is repeated as necessary. Applied exhaustively, this procedure generates the correct answer in all cases, yielding a classification of problems in terms of the number of models required to solve each one. However, different theories - and even different formulations of the same theory - disagree over the numbers in particular cases, such that the only reliable distinction is between syllogisms requiring just one model to solve correctly and those requiring more than one. It has been shown that multiple-model problems elicit more erroneous responses from participants than do single-model problems (Johnson-Laird and Bara, 1984; Johnson-Laird and Byrne, 1991).

Abstracting over the detail of both model- and rule-based theories, 'individual identification algorithms' (Stenning and Yule, 1997) yield another performance-related factor, this time influencing primarily the order of end terms in conclusions. While model-based theories seek to encapsulate whole states of affairs by positing as many entities as necessary, these algorithms focus explicitly on discovering a single, 'critical' individual that must exist, given the truth of both premises.[7] This reveals that in the great majority of cases the existence of the critical individual is entailed by just one of the premises. This is termed the 'source' premise, while the other is the 'conditional' premise. A full description of this individual is constructed using all three terms in the syllogism, with the two from the source premise coming first, in order, followed by the remaining one from the conditional. Stenning and Yule show that that the order of the end terms in such a description is an excellent predictor of the order of terms in conclusions drawn by participants. That is, according to their *Source Founding Hypothesis*, a conclusion will contain the end term from the source premise in the subject and the end term from the conditional premise in the predicate.[8]

[7] The reader is reminded of the 'no empty sets' assumption.

[8] Additionally, this framework identifies a distinct step from the individual description to a quantified statement in which much of a problem's difficulty might reside.

2.2.2.3 Individual differences

The different reasoning styles of 'visualisers' and 'verbalisers' lead to their experiencing different degrees of difficulty in solving the same syllogism (Ford, 1995). Moreover, the efficacy of logic instruction on them is significantly affected by the congruence or otherwise of its mode of delivery, graphical or sentential, with their preferred mode of reasoning (Monaghan and Stenning, 1998). However, Stenning and Yule (1997) present individual identification algorithms framed both diagrammatically and as a species of propositional calculus and argue that representational differences reside at a level of implementational detail too fine-grained for current theories to capture. Instead, Stenning and Cox (2006) suggest that Ford's results might be accounted for by reference to premise ordering. Updating the term 'canonical' to denote cases in which the source premise comes first, followed by the conditional, they observe that visualisers tend to perform better than verbalisers on non-canonical orderings.

Leaving preferences for different modalities aside, then, conclusion drawing has been shown to be influenced by preferences for different interpretations of the premises, as revealed by the *immediate inference* task. Here, participants are instructed to assume the truth of a single premise and then evaluate a second relating the same two terms. 'Hesitancy' is the tendency to respond CAN'T TELL when the correct answer is either TRUE or FALSE and 'rashness' is the tendency to respond either TRUE or FALSE when the correct answer is CAN'T TELL. These behaviours are exhibited variously by different groups of participants according to whether the order of the terms in the second premise is the same as in the first ('in-place') or reversed ('out-of-place'). Aside from the significant minority of participants who are neither rash nor hesitant at all, most are either rash in-place (RI) or hesitant out-of-place (HO) only. Smaller numbers are both RI and HO or RI and RO (rash out-of-place) (Stenning and Cox, 1995).

In the full reasoning task, these traits mediate the overall influence of several structural factors in a raft of complex interactions (Stenning *et al.*, 1996; Stenning and Cox, 2006). So, premise order is a strong influence on conclusion term order in all cases - that is, AC conclusions predominate - but especially so for HO participants. Likewise, the figural effect on conclusion term order is powerfully manifest across the board - end terms in conclusions generally retain the grammatical rôle s they have in the premises, where these differ - but again much more so for HO participants. Greater differences are linked to particular premises and quantifiers, notably the negative ones. In direct contrast to the other quantifiers, *No*, especially when in the second premise,

encourages participants to place the end term from its premise in the predicate of the conclusion. This time, though, the general effect is significantly less pronounced in HO participants than in others, while for RO participants it is reduced when *No* is in the first premise. For RI participants, on the other hand, *Some...not* in the first premise has an effect like that of *No*.[9] The figural effect interacts with some of these specific effects to enhance or, curiously, reverse them.

2.3 Discussion: interpretation and inference

Stenning and Cox's (2006) findings, summarised in the last part of the previous section, are given by way of a riposte to a 'scandalous' suggestion. That is that inference in the syllogistic reasoning task does not proceed on the back of interpretations of the individual premises, but instead follows an altogether different course from which they are invisible. This suggestion comes from investigations carried out by Newstead and colleagues into two types of supposed errors of misinterpretation (Newstead, 1989, 1995; Newstead and Griggs, 1999; Roberts *et al.*, 2001). These are errors of illicit conversion, which have already been described, and errors of 'implicature', which are embedded in Grice; Grice's (1975; 1989) highly influential account of the relationship of 'conversation' to logic.

Grice contends that ordinary, everyday conversation differs from the pursuit of logical deduction in that its purpose is to convey information held by one participant to another, who, it is assumed, does not currently hold that information. In other words, its aim is to transmit a single, coherent conception, or model, of the world or some state of affairs from the speaker to the hearer. This aim is held in common by the two participants and their shared, implicit understanding of it enables them to co-operate in its fulfilment. As a result, utterances are both generated and interpreted with respect to a set of shared conventions ('maxims') which streamline the process. They do this by licensing assumptions, called 'implicatures', additional to the actual propositional content of an utterance and not logically determined by it. They are derived instead from the choice of just that content and the particular form of its delivery, given the alternatives available in the context of utterance. In this way, a speaker can leave many details unstated for the hearer to fill in, resulting in the economical transmission of the intended model.

[9] Rashness in-place is also the only trait that has an impact on reasoning accuracy, tending to generate more invalid conclusions.

By contrast, the aim in logic is not to convey a particular model, but to demonstrate the logical dependence or independence of propositions under any possible interpretation of them. So, the assumptions of conversation are absent and the drawing of implicatures over and above the explicit propositional content of utterances is not licensed. Untrained individuals, it is claimed, make logical errors on account of their failure to relinquish these ingrained habits of interpretation. The standard example of this in the domain of syllogisms is the error of mistaking the quantifier *Some* to mean *Some but not all* rather than *Some and possibly all*. This is said to be due to the inappropriate application of what Grice termed the maxims of Quantity, which, in paraphrase, state that a speaker should give neither more nor less information than the situation demands. A hearer presented with *Some* assumes that it excludes *All* because if the speaker had known and intended *All* that is what he would have said instead. Consequently, given *Some A are B*, the hearer draws the implicature *Some A are not B*, and vice versa.

Newstead (1989) investigated this pair of errors and the two possible errors of illicit conversion, namely that *All A are B* entails *All B are A* and *Some A are not B* entails *Some B are not A*. Using two different tasks as diagnostics, he found that both types of misinterpretation of premise statements were indeed common. Implicature-style errors were revealed by requiring participants to indicate which Euler's Circle diagrams corresponded to each of several different premise statements, while conversion-style errors were manifested in an immediate inference task. Furthermore, the reasoning errors that would be expected as consequences of these interpretational errors were indeed committed in a syllogism solving task by the same participants who made the interpretational errors. The conclusion, however, was that conversion provided the better explanation overall, since it also accounted for the great majority of potentially Gricean reasoning errors that were committed, namely those involving *Some...not*. Errors involving *Some*, explicable solely in Gricean terms, were surprisingly rare in the reasoning task, given the results of the interpretation task.

This last, curious finding was pursued in a follow-up study (Newstead, 1995). There, several combinations of variants on both the interpretation and reasoning tasks were employed. Again, Gricean errors proved to be very common in the pure interpretation task, but less so when participants were required to generate their own inferences than when required to evaluate candidate inferences presented to them. Fewer still were made in the evaluative condition of the full syllogism task and hardly any in the conclusion-generation condition. From this it was concluded that the processes

of interpretation and reasoning are in fact so distinct that they invoke entirely different mechanisms, such that the former has no significant impact on the latter. It is this rather startling result that has been challenged by Stenning and Cox (2006).

They claim that this, along with much previous research into syllogistic reasoning performance, is flawed due to the combination of an inadequate conception of the task of deduction and an over-restrictive focus on specific implicatures and the potential errors associated with them. These failings derive, at least in part, from shortcomings in Grice's theory itself. This, it will be recalled, presented 'conversation', or exposition, as a form of discourse in which logic is augmented with pragmatics in order to achieve the goal of communicating information. In so doing, Grice cast logic as exposition stripped of pragmatics and communicative purpose. That is, perhaps inadvertently, he failed to acknowledge proof, or deduction, as a form of discourse also, intended to convey *something* and possessed of a pragmatics of its own. A pivotal element in this is the notion of logical independence.

A conclusion is said to be logically dependent upon one or more premises if its truth value is completely determined by the truth values of those premises under any interpretation of them. If there is at least one interpretation which leaves it open for the conclusion to be TRUE or FALSE, then it is logically independent of the premises. So, a grasp of deductive discourse is indicated by a willingness to conclude CAN'T TELL as distinct from either TRUE or FALSE when appropriate. Crucially, though, Newstead (1995) did not offer participants the option of responding CAN'T TELL in the immediate inference task. The effect of this is to conflate the responses FALSE and CAN'T TELL in opposition to the response TRUE. In obscuring this vital distinction, the test therefore failed to separate participants with an understanding of logical validity from those wedded to assumptions of informativity.

Reflecting on the fact that both TRUE and FALSE express logical dependence also leads to a broader conception of implicature than the traditional. Newstead (1995) adopts the orthodox view that responding TRUE to the conclusion *Some A are not B* when given the premise *Some A are B* is an error of Gricean implicature whereas responding FALSE is not. FALSE is, however, just as erroneous an answer as TRUE and can be arrived at by much the same sort of reasoning. Since *Some* is just as compatible with *All* as with *Some...not*, the speaker's failure to qualify his statement leaves the hearer at liberty to assume that the only *As* in the intended model are those that have been mentioned, i.e. that, within the local context of the discourse, *All A are B* is TRUE

and *Some A are not B* is FALSE.[10]

Another possibility is that the speaker elected to use *Some* as opposed to any other quantifier precisely because he was unsure of the truth or falsehood of any other assertion. The Gricean account assumes that the speaker is omniscient. Furthermore, as Stenning and Cox point out, Gricean theory does not even provide a principled basis for excluding implicatures in the reverse direction - *Some B are not A*, for example - which is in fact as good a definition of conversion error as any other. Consequently, limiting the investigation to just the paradigm cases is not sufficient to capture the full range of potential interpretational behaviour. Every possible combination of quantifiers and term orders in both the immediate inference and full reasoning tasks must be exploited in order to track down the sorts of implicatures that participants actually draw.

The critical point here is that the demands of informativity in 'conversation' only determine that implicatures, defined now more broadly as conclusions both TRUE and FALSE in either direction, may be drawn by the hearer, whereas in the absence of those demands he CAN'T TELL. What is missing from the theory is an account of how the hearer knows, or can be expected to know, which particular implicatures are warranted and which are not. Stenning and Cox argue that it is the linguistic devices of information structure, or 'information packaging' (Vallduví, 1992), that provide this detailed guidance. These are asymmetrical structures which signal what an utterance is intended to be informative about and what information it is intended to convey. In so doing, they circumscribe the range of alternatives from which the utterance was chosen.

Under this conception of the difference between expository and deductive discourse, the transition from everyday language use to full competence in logic requires the realisation of two distinct insights: not only that statements may be logically independent of each other, but also that the propositional content of an utterance is neither modified nor constrained by its linguistic structuring of the information it contains. This framework enables Stenning and Cox to characterise the differences between rashness and hesitancy, in- and out-of-place, on the assumption that these two points can be grasped at least semi-independently of each other and to different degrees by different individuals. Thus, one who is both RI and RO, although wedded to the idea of drawing implicatures, is not constrained by subject-predicate ordering and is therefore

[10] See Stenning (1996) for a more naturalistic discussion revolving around the example *Some boys are running* which makes obvious the introductory nature of this use of *Some*.

free to draw them in both directions. One who is RI only, on the other hand, while similarly unable to resist drawing implicatures, is guided by the linguistic structure and consequently restricted to classically 'Gricean' errors. In contrast, one who is HO perhaps does show some grasp of logical independence but remains sufficiently influenced by subject-predicate ordering to prevent them drawing some out-of-place conclusions that are in fact valid. One who is both HO and RI, though, is so influenced by it that it overwhelms any other factors.

The flaw in this argument, however, is that theories of information structure in fact advert to subject-predicate ordering very little (see Chapter 3). That distinction is primarily a grammatical one, as Stenning and Cox's own use of the alternative term "grammar" for it acknowledges. Now, syntactic alternatives do provide some ability to tailor the presentation of propositional content, but even Vallduví's (1992) heavily syntactically influenced treatment deals more with left- and right-dislocation than with subject and predicate. The primary vehicle for conveying information structure, in English at any rate, is not word order but prosody, especially intonation. So, if Stenning and Cox's underlying theory is right, the articulation of statements signalled by intonation should feature as a factor in the performance of syllogistic tasks. [11]

Once attention is turned to information structure as manifested in prosody, it may immediately be noted that this element of the linguistic signal is not apparent in written materials. This is not to say that they lack information structure; rather, it is for the reader to recover it from context. Syllogisms, however, provide minimal context and participants are therefore at liberty to impose on them whatever structures make sense to them. Hitherto, this aspect of the task has neither been controlled for nor investigated, so the strategies that participants use to assign information structure to premises in syllogisms are simply not known. This, then, is the basis for the series of experiments that form the core of the present work.

[11] Indeed, although their analysis is cast entirely adequately in terms of the interaction of two variables, it is nonetheless dissatisfying to have the explanation for rashness apparently based on adherence to information structure when in one direction but on disregard for it when in the other. Similarly, it is odd to find RI+RO and RI+HO to be near opposites when they share the single most prevalent trait in common. In English, the at least bi-directional articulations of information structure have a clear advantage over the uni-directional one of subject-predicate ordering in the search for a more uniform treatment.

Chapter 3

Information structure in syllogisms

3.1 Theories of information structure

3.1.1 Introduction

Information structure is one of the less well-defined components of linguistic systems and this is reflected in the profusion and confusion of terms that have been coined to describe it. However, as this lack of consensus has become an increasingly apparent impediment to progress, so commentaries on it have at least begun to draw out the parameters of agreement and disagreement, as shall be seen. Crudely put, information structure concerns what an utterance is 'about' and, for this purpose, identifies some part of it as having some rôle distinct from the rest. The bipartite, asymmetrical articulations that result from this are generally recognised as themselves falling into two categories.

First are those that distinguish the part of an utterance that identifies the matter with which it is concerned from the remainder, which conveys something concerning that matter. In this tradition, Halliday (1967) calls the *theme* that which acts as the "point of departure for the clause as a message" and the *rheme* the thrust of the message. The examples in 3.1 (from Halliday, 1967) show three different assignments of theme (in bold) and rheme (the remainder) for the same proposition.

(3.1) a. **John** saw the play yesterday.

b. **Yesterday** John saw the play.

c. **The play** John saw yesterday.

Many writers prefer the term *topic* over theme, its complement being usually the *com-*

ment rather than the rheme. (e.g. Hockett, 1958; Gundel, 1974, 1988; Schmerling, 1976). So, for example, Hockett observes straightforwardly that topic and comment are suggestive of "the most general characteristic of predicative constructions ...: the speaker announces a topic and then says something about it." The majority, however, do not insist on the topic being the leftmost constituent in every case, especially in spoken English. Thus, Vallduví (1992) notes that both 3.2 and 3.3 are entirely acceptable.

(3.2) a. Where can I find the flatware?
The forks are in the cupboard...

b. but **the knives** I left in the DRAWER.

(3.3) a. Where can I find the flatware?
The forks are in the cupboard...

b. but I left **the knives** in the DRAWER.

The second category contains those distinctions based on the relative prominence of the elements in an utterance, such that one is 'foregrounded' as against the rest. In English, with its fairly constrained word order, this is achieved by prosodic accentuation of the foreground constituent (Ladd, 1996).[1] The examples in 3.4 show the same proposition worded the same way but with three different accenting patterns. (Accented words are indicated by SMALL CAPITALS.)

(3.4) a. The dog bit the BOY.

b. The DOG bit the boy.

c. The dog BIT the boy.

While 3.4a is, all other things being equal, the most naturally narrative-sounding of the three, i.e. appropriate in actual or putative response to the open question "What happened next?" they are each easily imagined as answers to just one of the following three specific questions about different aspects of the same event.

(3.5) a. Whom did the dog bite?

b. What bit the boy?

c. What did the dog do to the boy?

[1] In languages with freer syntax it is typically done by moving the constituent to a prominent position in a more fixed intonation contour. See Vallduví (1992) for numerous cross-linguistic examples.

The accented element is commonly referred to as the *focus* of the utterance, while the remainder has received a variety of names, including *ground* (Vallduví, 1992), *background* (Dahl, 1974; Chafe, 1976), *presupposition* (Chomsky, 1971; Jackendoff, 1972), and *open-proposition* (Prince, 1986). The distinction between the 'open' and specific uses of 3.4a is often cast as being between 'broad' focus (Ladd, 1996) and 'narrow' focus (Selkirk, 1984), which attests to the fact that, although first and foremost attributes of individual words, accents can in some sense be associated with larger constituents.

Instead of focus and ground, Halliday (1967) used the terms *new* and *given*, respectively, in an attempt to highlight what the distinction is believed to signify, which is that the former contains what is genuinely informative in the utterance while the latter provides the setting for that information. However, it has long been clear that neither term should be taken literally. On the one hand, something may be 'given' in virtue of implication or general background knowledge rather than explicit prior mention. On the other hand, something may be 'new', i.e. informative, despite having been recently mentioned, in virtue of there being other possible occupants of its position at that stage of the discourse.

Now, evidently these two characterisations of 'aboutness' can be conceived of simply as representing the same thing seen from opposing standpoints, with the difference lying in the object of primary interest. Indeed, in very simple statements the two coincide exactly, presenting the positive and negative images of the same photograph, as it were. This is typically seen in cases of broad focus, such as 3.6, adapted from Vallduví (1992). (Topic is indicated by $_T$, comment by $_C$, ground by $_G$, and focus by $_F$.)

(3.6) a. Tell me about John. What does he do?

b. $[_T$ **John**$]$ $[_C$ drinks BEER$]$.

c. $[_G$ **John**$]$ $[_F$ drinks BEER$]$.

In the majority of cases, however, there are spans which belong neither to the topic, on the one hand, nor to the focus, on the other, giving rise to a dynamic tension in the overlap between ground and comment. This is illustrated in the narrow focus variant of 3.6, adapted from Vallduví (1992) and Dahl (1974), shown in 3.7.

(3.7) a. Tell me about John. What does he drink?

b. $[_T$ **John**$]$ $[_C$ drinks BEER$]$.

c. $[_G$ **John** drinks$]$ $[_F$BEER$]$.

This dynamic is whole-heartedly embraced in the Functional Sentence Perspective of the Prague School (Sgall, 1967; Sgall *et al.*, 1986; Firbas, 1992), which adopts a view of the sentence as a vehicle for getting from one pole to another. It is unsurprising, then, that topic and focus, each the most commonly used term in its category, are the names it employs for those poles. A contrastingly reductive resolution of the conflict is offered by Vallduví (1992). This conflates the two distinctions into a single, tripartite articulation in which the ground is subdivided into a *link* (essentially equivalent to a topic) and an optional *tail*, each of the three constituents being functionally distinct. Thus, 3.6 and 3.7 become 3.8 and 3.9, respectively (Vallduví, 1992, adapted, with $_L$ indicating link and $_t$ tail).

(3.8) a. Tell me about John. What does he do?

b. $[_G\ [_L$ **John**$]]\ [_F$ drinks BEER$]$.

(3.9) a. Tell me about John. What does he drink?

b. $[_G\ [_L$ **John**$]\ [_t$ drinks$]]\ [_F$BEER$]$.

Explicitly hybrid approaches are rare, however, as the difference between the two viewpoints has, in truth, as much to do with the nature of the enquiry as with its object. All too common, though, has been blurring of the distinction, as scholars have grappled with their intuitions regarding what is admittedly a slippery notion, resulting in not merely a confusion of terms but more of a continuum of theories than the dual classification suggests. The Prague School on its own is home to multiple variants, some pursuing a 'combining approach' that conflates theme with given while others adopt a 'separating approach' which draws out differences between the two (Fries, 1983). Moreover, with particular reference to English, the further separation of theme from topic has been suggested, the former being the strictly positional starting point of the message, as in Halliday's (1967) scheme, and the latter the substance of what the message is about (Downing, 1991). From this it can be seen that the underlying dichotomy that is and has been all too easy to lose sight of is essentially that of the relationship between semantics and syntax, on the one hand, and semantics and pragmatics, on the other. To put it another way, theories of information structure are all attempts to understand the relationship between constituency and context, but such attempts are prone to emphasising one over the other.

3.1.2 Constituency versus context

Constituency-oriented approaches are concerned with the internal organisation of an utterance as a system, its parts working together to achieve the communicative purpose of the whole.[2] The parts are therefore not merely complementary but functionally interdependent and defined by their relationships to each other. Context-oriented approaches, on the other hand, are concerned with those connections which locate an utterance, or, rather, elements of it, within the wider communicative situation. This is a more open-ended enquiry in which those elements are related not to each other but to further elements outside the utterance in some over-arching model, be it of the text or discourse, the cognitive states of the participants, or the world.

Now, the first observation to make about these two kinds of approaches is that only constituency-oriented ones have a particular association with analysis at the level of the clause. Context-oriented approaches are really concerned with the status of sub-clausal elements, i.e. words or phrases. Unfortunately, the highlighting of one or more of them produces a pattern of 'foregrounding' and 'backgrounding' that is all too tempting to construe as an articulation of the clause. To do so, however, is to mistake variations in discrimination for relations of import, as the contrasts expressed thereby are not between neighbours in the utterance, but between elements in the larger model which are alternative candidates for occupying each particular slot in the utterance.

What is a more obvious point to make is in fact also a corollary to the first one. The constituency-oriented approach is not properly concerned with context. It identifies constituents simply by reference to the rôles they play in the delivery of the clause's message. This inevitably casts one of them as being in some way 'prior' to the rest. Again unfortunately, this insinuates the idea that it is 'prior' to the clause or message itself and therefore part of the context. This is a misapplication of the rôle outwith the scope of its definition, which is internal to the clause, leading to confusion of the slot with the filler.

There are other ramifications to this distinction, but at this juncture it is perhaps best to illustrate the differences by reference to other formulations. Three competing characterisations have been chosen for this purpose. They are Gundel's (1999) division of

[2]For most treatments, 'utterance' here can be replaced with 'statement' and 'communicative' with 'informative', but broader coverage is possible. For example, Gundel (1988) presents the following definition of topic: "An entity, E, is the topic of a sentence, S, iff in using S the speaker intends to increase the addressee's knowledge about, request information about, or otherwise get the addressee to act with respect to E." For the purposes of the present work, however, the narrower conception is satisfactory.

given and new into two types, 'referential' and 'relational', Vallduví's (1992) comparison of "topic-comment" and "focus-ground", and Kruijff-Korbayová and Steedman's (2003) distinction between "topic/comment" or "theme/rheme" on the one hand and "background/kontrast" or "given/new" on the other.

Gundel (1999) presents two types of given and new framed in terms very close to the distinction drawn here. The referential type "describes the status of an expression vis-a-vis a model of the world, the discourse, or the speaker/hearer's mind", while the relational type concerns "two complementary parts, X and Y, of a linguistic presentation, where X is given in relation to Y, and Y is new in relation to X." The main goal of this characterisation is to distinguish pragmatically or cognitively oriented analyses from 'linguistic', 'grammatical', or structural ones. These are obvious matches to the orientations of the two approaches as described above. However, the idea that both are variations on the articulation of given and new leads to a conflation of the two approaches, importing referential attributes to elements on the basis of relational status. This is illustrated by the use of the following example (originally from Gundel, 1980) to demonstrate the logical independence of the two types. The claim is that the pitch accent on SHE arises from the relational newness of the embedded subject with respect to its topicalised predicate, in spite of the presumed referent's undoubted givenness.

(3.10) Who called?
Pat said SHE called.

This is essentially the error of foregrounding/backgrounding producing articulation that was mentioned above. The question sets up a situation containing a default predicate with an open-ended subject. In the answer, the subject's 'newness' is due to its contrast not with the predicate but with other putative subjects that are contextually available. That is, it is new in that position independently of the predicate's being given in its position and irrespective of the relationship between the two. In 3.11, the subject is still accented, even though the predicate is also, whereas the 'relationally deaccented' subject in 3.12 carries with it a note of contradiction that is not warranted:

(3.11) Who drove?
Well, Pat said SHE WALKED.

(3.12) Who drove?
?Well, Pat said she WALKED.

Consistent with this assimilation of contextual elements to constituents is the inclusion of focus/presupposition (Chomsky, 1971) in the relational category. The more

conventional view places it alongside focus/ground rather than topic/comment. Treatments oriented around propositional content, though, as in fact both Chomsky's (1971) and Gundel's (1999) are, necessarily fixate on the clause, viewing one constituent as an instantiated positional variable that completes the proposition. It is symptomatic of them to privilege the focus as a special case of accenting attached to this constituent.

With respect to topics, though, Gundel avoids the error of 'priority' mentioned above by distinguishing between a 'syntactic topic' and a 'semantic or pragmatic topic'. This enables her to account for examples of non-referential topics such as the left-dislocated phrases in 3.13 and 3.14, taken from **?**.

(3.13) Most middle-class Americans, when they look at the costs plus the benefits, they're going to be much better off.

(3.14) Any company, if they're worth 150 million dollars, you don't need to think of ....

Noting that phrases such as these are pronounced with stress on the quantifier and that they typically have a partitive reading, she argues that the quantifier is part of the syntactic topic but not part of the semantic/pragmatic one. 3.13 and 3.14 can therefore be paraphrased as 3.15 and 3.16,respectively.

(3.15) (As for) Middle-class Americans, when most of them look at the costs plus the benefits, they're going to be much better off.

(3.16) (As for) companies, if any one of them is worth 150 million dollars, you don't need to think of ....

In this way, the constituent and the contextual components of information structure are shown to cross-cut each other, not be coextensive, with referential givenness and newness being attached to the latter and expressed through intonational variation.

In Vallduví's (1992) analysis, the two dimensions are certainly not coextensive, yet they are both firmly tied to clausal constituency. The tripartite articulation that this produces has already been mentioned, but it is the characterisations of both focus/ground and topic/comment underlying it that are of primary interest here. Not unlike Gundel's, Vallduví's interest resides essentially in propositional content and its delivery, leading to the near-complete binding of focus/ground to constituent structure. However, as the distinction between this and topic/comment is rigorously maintained, this imports the contextual aspect of the one into the other in quite a fragmentary way. In so doing, this

heavily constituency-oriented approach diverges instructively from Gundel's (1999) conception of new/given, particularly the referential variety, and highlights a genuine difficulty in respect of constituency-based analyses.

That Chomsky's focus/presupposition is now to be found in the focus/ground category is a minor, if telling, point. More significantly, explicitly excluded from it are notions of focus concerned with the referential status of discourse entities. Characterisations of the salience of entities, such as 'identifiability' (Chafe, 1987) and 'activation' (Lambrecht, 1994), are at the heart of Gundel's referential given/new, but in Vallduví's view they are precluded from playing any part at all in information packaging, because they apply to elements below the level of the clause. By much the same token, prominence due to accenting is considered to be a heterogeneous phenomenon, only some of whose instantiations are expressions of information structure. Specifically, the accent that identifies the focus is the only necessarily informational one, while there is a 'tune' which is a sufficient, but not necessary, topic marker. Any other accenting, or deaccenting, of elements in an utterance must, perforce, be unrelated to information structure.

Illustrative of this exclusionary stance is the discounting of some varieties of narrow focus, such as in 3.17a, adapted by Vallduví from van Deemter (1992). That this is indeed a case of narrow focus is attested by Ladd's (1996) example of it, shown in 3.17b (with SMALL CAPS added).

(3.17) a. The men in the hospital looked horrible. Especially the OLD men.

b. I didn't give him three francs, I gave him FIVE francs.

For **?**, the informationally relevant analysis stops at observing that the focus phrase as a whole ('the old men' in his example; 'five francs' in Ladd's) is realised with a focal pitch accent on one of its elements. In spite of the recognition that it signals givenness and newness, the fact that this accent is not on the phrase-final head noun is considered a deaccenting phenomenon driven by pragmatic factors concerning referential status solely at the phrase level. It is therefore totally unconnected with the specifically clause-level packaging of a proposition. This is to be distinguished from cases in which the focal accent is displaced from one phrase to another, as in 3.18, also from Vallduví (1992), which does have structural consequences. In 3.18a, the phrase 'on the table' constitutes all or part of the focus, but in 3.18b it is part of the ground (the *tail*, in Vallduví's terminology).

(3.18) a. John $[_F$ left a note $[_F$ on the TABLE$]]$.

b. John $[_F$ left $[_F$ a NOTE$]]$ on the table.

Tellingly, though, where there is ambiguity between broad and narrow focus, as illustrated here by the multiple bracketings of both strings, the extent of the focal constituent is determined, according to Vallduví, by reference to the context.

The lack of any necessary marking of topics is, as Vallduví recognises, a point of weakness for topic/comment theories in general. Topics are in fact claimed to exist on a variety of grounds, including accenting, deaccenting, dislocation, and contextual status. Given this apparent miscellany, there is arguably no means of identifying topics *qua* topics reliably and consistently in every case. Theoretical responses to this difficulty take one of two basic forms. One is to admit the possibility of topicless constructions, such that a whole statement may be a comment not directed to any particular thing. The absence of the topic marking(s) of choice in any given case is then unproblematic. The other is to adopt a strictly positional approach, insisting that the topic is always the leftmost, say, constituent in an utterance, dislocated or not. This guarantees an unequivocal identification in every case, albeit not infrequently of a constituent completely unmarked in virtually every other sense. On this point, Vallduví notes that there is in fact a degree of congruence between topics and grammatical subjects that has not gone unnoticed before (van Oosten, 1986).

Vallduví opts to admit the possibility of topicless constructions, but also, surprisingly, allows topics to be realised in a variety of ways. It seems that a 'link', as it is here termed, can be realised by any one of

1. a distinctive accent by itself. (e.g. as in 3.3, earlier.)
2. a distinctive accent plus left dislocation. (e.g. as in 3.2, earlier.)
3. possibly left dislocation by itself. (Left as an open research question.)
4. No overt marking at all. (Used with non-contrastive subjects.)

In this last case, the appeal is again made to context in order to determine whether there is a topic or not. In fact, though, it frequently appears to be context that drives the determination, even when there is overt marking. For examples, the link interpretation of 'the knives' in 3.3 is said to be 'forced' on it by the context.

This multiplicity of realisations appears to be motivated partly by cross-linguistic factors. However, the recurring significance of context suggests that it is primarily

necessitated by the particular semantics adopted. Following Heim (1982), information is said to be stored on 'file cards', each entity introduced into the discourse having one. It is the function of a link to identify an existing card as the locus of information update. There is, then, an imperative in many instances for there to be a link.

As its name and hybrid nature imply, the rôle of a link in Vallduví's scheme is to 'anchor' an utterance in the prior discourse. This is a common conception of topics, shared not only with Gundel (1999) but also with Kruijff-Korbayová and Steedman (2003), who distinguish between "the part of an utterance that relates it to the discourse purpose, and the part that advances the discourse." However, this view of theme and rheme, as they term them, is contrasted with the explicitly contextually-oriented distinction between *kontrast* and *background* (Vallduví and Vilkuna, 1998), which concerns "parts of the utterance - actually, words - which contribute to distinguishing its actual content from alternatives the context makes available." These two dichotomies are thus conceived of as dimensions that cross-cut each other, introducing an element of discontinuity into constituents. This is illustrated in 3.19, which is adapted from Steedman (2000).

(3.19) Q. I know that Marcel likes the man who wrote the MUSICAL.
But who does he ADMIRE?

A. (Marcel ADMIRES) (the woman who DIRECTED the musical)
Background *Kontrast* *Background* *Kontrast* *Background*
Theme *Rheme*

Steedman (2000) keys this directly to a version of 'Alternative Semantics' (Karttunen, 1976; Karttunen and Peters, 1979; Rooth, 1992; Büring, 1997) wherein the context contains two sets of candidates for the content of a given constituent, one each for theme and rheme. This makes it clear that the kontrast/background distinction is not so much a relational one between neighbouring constituents as a discriminatory one between the salient features of competitors for a given slot. For Kruijff-Korbayová and Steedman, then, the notion of theme is free to be considered "tightly related" to the "notion of a center of attention" (Grosz *et al.*, 1995; Walker *et al.*, 1998), whereas in Vallduví's view the latter concerns the referential status of sub-clausal entities and is therefore completely unrelated to information structure.

This more decoupled characterisation has the flexibility for other schemes to be cast straightforwardly into its terms. For example, broad focus effectively corresponds to the combination of rheme and kontrast, while narrow focus is more specifically kontrast found within rheme. Kontrast within theme, on the other hand, is comparable to

Vallduví's (1992) overtly signalled link. The framework also enables broader perspectives to be taken on issues such as the rôle of pitch accents and the number and nature of the levels of structure involved.

Kruijff-Korbayová and Steedman distinguish between theories in the tradition of Bolinger (1965) that give all pitch accents fundamentally the same interpretation and those, notably of the Prague School, that consider there to be different kinds of accents specific to the constituents to which they are attached. The former take the categorically context-oriented view that accentuation expresses the same fundamentally contrastive idea wherever it appears. Vallduví and Vilkuna's (1998) terms, kontrast and background, actually belong to this camp. More constituency-influenced approaches of the latter kind, on the other hand, effectively limit this to kontrast within theme, treating kontrast within rheme as a distinct and non-contrastive expression of focus. (Additionally, these should in turn be distinguished from theories at the even more constituency-bound end of the scale, such as Vallduví's (1992), which separate out the functionality of accents other than the focal from information structure altogether.)

As for the possibility that there is more than one level of information structure, in fact the only case that Kruijff-Korbayová and Steedman cite is that of Halliday (1967), whose conceptions of theme/rheme and given/new are also fairly distinctly constituency- and context-oriented, respectively. Indeed, the only real drawback of viewing them as levels rather than dimensions is that it does not allow the interaction of the two to be adequately addressed. The great majority of theories, as has been seen, do not distinguish clearly between the two dimensions and therefore take them to exist at one and the same level of structure. This also gives rise to both the possibility of recursion and its denial. Kruijff-Korbayová and Steedman note that, while most analyses admit mild forms of recursion, usually involving co-ordination or subordination, extreme positions are also possible. Clearly, strongly propositionally-oriented approaches can allow no recursion, as they allow no information structure below the clause level. Intermediate positions that merge constituency with foregrounding and backgrounding, on the other hand, have the capacity to view utterances as 'Russian dolls' of nested structures.

Of the three very different conceptions of the dichotomy at the heart of information structure, then, this last comes closest to embodying that between constituency and context. Even it does not succeed entirely, though. Indeed, relating theme not just to the context but to the discourse purpose is arguably a greater merging of the two orientations than occurs in either of the other two. However, it is also perhaps more of

a pragmatic connection than a definitional one, as it in all likelihood comes from the speaker-oriented viewpoint of Steedman (2000). This provides an alternative means of identifying a theme to the purely positional one considered by Vallduví which is yet constituency-based. It is that what is theme and what is rheme is decided by the speaker and signalled by them by means of phrasing. Now, in practice, a speaker is likely to adopt a discourse-level theme as an utterance-level theme. This is not unlike Gundel's empirical connection between relationally and referentially given, but weaker in that it concerns the maintenance of coherence relations across the discourse. At the utterance level, though, this is achieved through the felicitous handling of the contextual component so as to establish a referent in common with the themes and/or rhemes of other utterances. That is, discourse dynamics arise out of the interplay between the two distinctions, not just one of them.

Turning, finally, to present purposes, though, the speaker-based approach is not appropriate, as the intention is to compare a speaker-independent analysis of syllogisms as minimal texts with the productions of a number of different speakers. Of the text-based theories, Halliday's (1967) clause-initial view of theme is the most purely constituency-oriented. Given the restricted nature of the constructions under investigation, however, there is no possibility of this identifying as theme anything other than the grammatical subject. Not only are subject and predicate less ambiguous terms than theme and rheme, they also enable a clearer connection with Stenning and Cox (2006) to be made. Staying with Halliday, new and given will be used for context-oriented description. However, this must be informed b y the insights that the fundamental concept underlying the is one of contrast and that this constitutes a dimension of analysis intersecting with that of theme/rheme rather than a separate layer parallel to it.

The two dimensions having been independently covered in this way, the terms topic and focus are free to be used in relation to certain intersections of them, much as they can be located within Kruijff-Korbayová and Steedman's framework. For one thing, this permits a topic to be considered either marked or unmarked, according to whether it is accented or deaccented, respectively. In thus giving meaning to phrases such as 'subject-focussing' and 'predicate-topicalisation', this arrangement of terms affords a degree of descriptive flexibility appropriate to the exploratory nature of the study at hand.

3.1.3 Predication and prosody

With the foregoing definitional matters settled, attention may now turn to the prosodic realisation of information structure in predicative statements[3], such as occur in syllogisms. In English, as has been seen, given or repeated items are normally deaccented, while new or contrasting items are marked with pitch accents. The main stress in a a sentence, marking its focus, is generally carried by arguments rather than predicates. In short, simple statements where the predicate is an intransitive verb, this tends to produce subject-focussing. This occurs especially if the statement serves to introduce the subject into the discourse:

(3.20) a. My KNEE hurts.

b. Your BROTHER called.

c. The SUN came out.

d. Her CAR broke down.

Where, however, the subject is given and can be readily recovered from the context, it is likely to be deaccented as the topic and the predicate focussed instead. This is Schmerling's (1976) explanation for her well known pair of examples, each reporting the death of a former president of the US at the start of a conversation, the first of which was expected but the second not:

(3.21) a. Truman DIED.

b. JOHNSON died.

Predicate-focussing also tends to happen when statements deal with generic subjects, definitions, generalisations, abstractions, and absolutes:

(3.22) a. Dogs BARK.

b. Toys BREAK.

c. The sun SHINES.

d. Money TALKS.

Gussenhoven (1983) terms the subject-focussed cases 'eventive' readings and the predicate-focussed ones 'definitional' or 'contingency' readings. The former are considered to

[3]This section draws significantly on Ladd (1996). As there, CAPITALS are used informally to signal accenting.

deliver unitary items of entirely new information ('news'), whereas the latter single out their subjects as already contextualised and distinct from the truly informative elements that are the predicates. Ladd (1996) observes that contingency sentences in fact often have a secondary accent on the subject argument. This seems to have the effect of introducing the subject, but only to serve as the locus of the real import of the statement, which is delivered by the predicate.

Contingency readings, then, are paradigm examples of topic and comment, in which newness in the predicate is focussed. Gussenhoven's conception of them is the straightforward one of an unmarked topic, produced by givenness in the subject, whereas Ladd's observation highlights the interaction of constituency and context more fully. There, the topic is marked because, although it is given, it is new *in subject position* in the sentence. That is, even though it is in some sense contextualised, it still requires to be picked out as the subject from among other possible subjects.[4] In eventive readings, by contrast, both subject and predicate are completely new. When the subject contains the only argument, that attracts the focus and the predicate is left unmarked.

Where the predicate contains a transitive verb and, consequently, a second argument, the difference in realisation between the two types of readings becomes more subtle. All other things being equal, it is the object argument that is the single most informative element, because, even if both arguments are new, the one in the subject is part of the 'point of departure' of the clause whereas the one in the predicate is part of its message. Therefore, focus is drawn to the predicate in eventive as well as contingency readings. The difference between the now unmarked subject and that of a Gussenhoven-style contingency reading then comes down to phrasing alone. That is, of course, unless the subject is secondarily accented, which remains likely in many cases. This is illustrated in 3.23, where 3.23a is a repeat of 3.4a and 3.23b is its secondarily accented equivalent.

(3.23) a. The dog bit the BOY.

b. The DOG bit the BOY.

Also possible, though, is an unmarked topic in the predicate. This occurs when the object argument is in fact given and therefore deacccented. Focus then reverts to the subject, or, less commonly, the verb. These correspond to the other two narrow focus variations on 'the dog bit the boy' shown in 3.4.

[4] See also example 3.1.2, earlier.

Still more marked accenting patterns arise in phrases containing combinations of given and new elements that serve to distinguish topics from other entities which are not only salient but also similar. In these cases, the degree of contrast produces an exaggerated, low-high 'tune', as illustrated by the following examples taken from Vallduví (1992) and Steedman (1991)[5] [6]:

(3.24) a. S_1: I know Mary's undergraduate degree is in physics, but what subject is her doctorate in?
S_2: Mary's **doctorate** is in CHEMISTRY.

b. S_1: Oh, I didn't know both of them have a doctorate...
S_2: Yes, but **Mary's** doctorate is in CHEMISTRY and **Anna's** doctorate is in LAW.

As touched upon earlier (see example b) Ladd (1996) uses the phrase 'five francs' to illustrate similar effects occurring in foci. Stress on the first element (FIVE *francs*) clearly signals narrow focus on the amount in a context where the currency can be taken for granted. Stress on the second (*four* FRANCS), however, might signal narrow focus on the currency specifically or broad focus on the sum of money as a whole. In fact, much the same ambiguity can occur in introductory, as opposed to contrastive, topics.

These, then, are the intonational building blocks of the articulations to be expected of the sort of predications found in syllogisms, on the assumption that such discourse is not so unnatural as to distort them significantly. The combination of generic quantifiers and situationally embedded terms, the latter being effectively necessitated by the 'no empty sets' assumption, makes the use of secondarily accented, introductory contingency readings seem likely, at least in first premises, where there is as yet no context beyond the introductory rubric. That is, if the situation is said to involve people of different nationalities and occupations, the first premise *Some Hungarians are busdrivers* is likely to be realised with accents on both *Hungarians* and *busdrivers* in a pattern similar to that in 3.23b.

In second premises, by contrast, a variety of given/new patterns is possible, arising out of the potential for repetition of quantifiers, the guaranteed reuse of the middle

[5] Steedman's view is that this tune actually identifies all true topics, or themes. However, in less contrastive contexts it is often very difficult to distinguish it from a standard high accent, as Steedman himself acknowledges (Steedman, 2004). This is especially the case with sentence-initial topics, which, as has been explained, are common.

[6] Reformatted slightly, with some detail omitted, for consistency and clarity

term, the strong syntactic parallelism of the premises, and the bipolar nature of the predications involved. Therefore, the full range of subject- and predicate-focussed patterns, featuring both marked and unmarked topics, may be expected.[7]

3.2 Information structure in exposition and deduction

3.2.1 Contextually concordant and contextually neutral intonation

It is now possible to relate credulity and skepticism more closely to information structure and intonation. It will be recalled that the essential difference between them is the presence or absence of the assumption of logical dependence, i.e. that successive statements are in some way linked to previous ones. This is a difference over the applicability of the contextual component of information structure. If exposition allows given and new to function as normal, then both subject and predicate can be topicalised or focussed. Credulous processing therefore has the capacity to generate rash implicatures both in- and out-of-place. If, in contrast, deduction discounts context, then only theme-rheme structure remains to indicate that what a statement is 'about' is its subject. Skepticism, therefore, limits the potential for hesitancy over inference to out-of-place instances only.

Figure 3.1 illustrates the idea in relation to the explicit contextualisation of second premises. Obviously, in the immediate inference task, participants are free to impose and, indeed, vary the information structure of statements as they see fit. In 3.1(i), credulous processing gives rise to the possibility of a rash in-place implicature relating to the given subject term (*Koreans*), whereas in 3.1(ii) the potential is for a rash out-of-place implicature concerning the given predicate term (*nurses*). Ignoring givenness and newness, on the other hand, skeptical processing restricts the scope of hesitant inference to the subject term and the consequent blocking of inferences relating to the predicate term (*nurses*) in both cases.

Following on from this, then, it is possible to posit two distinct sets of intonation contours that reflect the two different kinds of interpretation. The first set contains those variants which respect patterns of given and new and the second those that ignore them. For present purposes, the former shall be termed 'contextually concordant' (*CC*,

[7] Of course, it is possible that forward scanning may induce these more marked patterns in first premises also, in anticipation of continuation or contrast to come, but the impulse is not as strong. Second premises lacking such emphasis where it is warranted sound definitely odd and, for this speaker at least, are difficult to produce, whereas this is not true for first premises.

		(i) Figure BA-BC	(ii) Figure AB-CB
e.g.		*All Koreans are golfers* *Some (Koreans) are nurses*	*All golfers are nurses* *Some Koreans are (nurses)*
Rashness	$\Rightarrow$	[*Some (Koreans) are not nurses*]	[*Some (nurses) are not Koreans*]
Hesitancy	$\not\Rightarrow$	[*Some nurses are Koreans*]	[*Some nurses are Koreans*]

Figure 3.1: Topic-based rashness and subject-based hesitancy. Second premise subject term is shown in San serif font. Given term is shown in rounded brackets.

or simply 'concordant', for short) and the latter 'contextually neutral' (*CN*, or just 'neutral'). Now, obviously the two resolve to the same thing in the case of discourse-initial statements, which the prompts in the immediate inference task and the first premises of syllogisms may be taken to be. The real differences emerge in second premises, where the particular combinations of mood and figure generate a variety of given/new patterns. Pre-empting the work of the next section a little, this is illustrated informally in Figure 3.2 using the example syllogisms from Figure 3.1.

	(i) Figure BA-BC	(ii) Figure AB-CB
	All KOREANS are GOLFERS	All GOLFERS are NURSES
CC:	SOME Koreans are NURSES	Some KOREANS are nurses
	All KOREANS are GOLFERS	All GOLFERS are NURSES
CN:	Some KOREANS are NURSES	Some KOREANS are NURSES

Figure 3.2: *CC* and *CN* intonation of two sample syllogisms.

It follows from this, though, that the immediate inference task, even as improved by Stenning and Cox, can not lay these behaviours fully bare. Founded as it is on the interpretation of premises in isolation, it fails to control for the contextual component and over-emphasises syntax. This might, indeed, be the reason why the primary manifestation of the various behaviours in the full reasoning task is in the surface ordering of conclusions - term, subject-predicate, and theme-rheme ordering are all equivalent here, after all - rather than more substantively in whether or not conclusions are actu-

ally drawn, or whether or not they are logically correct.[8]

Fortunately, this revised account also makes feasible the observation of interpretational behaviour 'in situ' in the full reasoning task, where the combinations of mood and figure do give control over patterns of given and new, instead of indirectly via the immediate inference task. The principles detailed in the previous section can be used to create projections of the expected contents of both the *CC* and *CN* classes of intonation contours, allowing them to be empirically verified and compared more directly to performance in the full task. It should be obvious that the *CN* class in fact contains only one member which serves in all contexts and that this is identical to the first premise instance in the *CC* class. The next subsection therefore concentrates on describing the derivation of second premise *CC* contours.

3.2.2 Prediction of concordant intonation patterns

3.2.2.1 Figure

Table 3.1 profiles the terms in each of the four figures according to their givenness and newness, both independently and in conjunction with subject and predicate. From this it can be seen that figure AB-BC represents a simple topic shift from the first premise and so conforms to Ladd's (1996) conception of a contingency reading. The symmetrical figures, occupying the lower half of the table, are straightforwardly contrastive, featuring topics unmarked due to simple repetition. Figure BA-CB is the most unusual, as it features the introduction of a completely new topic followed by focussing of the previous one.

Thus, it is to be expected that figure AB-BC will have a secondarily accented subject term and a focally accented predicate term. In the two symmetrical figures, the middle term will be completely deaccented, with the other therefore receiving the focal accent. The subject term in figure BA-CB will receive an exaggerated accent owing to highly contrastive nature, while the predicate is focally accented. In all bar AB-BC, there is some degree of contrast expressed in the overall pattern of given and new which might be manifested in a degree of exaggeration of the focal accent, but this will not be assumed here.

[8] Perhaps it also explains the somewhat anomalous combination of RI+HO, or even RI without RO.

Subject term	Predicate term
A	B
B	C
given	*new*
new as subject	*new* as predicate
B	A
C	B
new	*given*
new as subject	*new* as predicate
A	B
C	B
new	*given*
new as subject	*given* as predicate
B	A
B	C
given	*new*
given as subject	*new* as predicate

Table 3.1: Givenness and newness of second premise subject and predicate terms in the four figures.

3.2.2.2 Mood

The combination of elements embodied in mood requires a more sophisticated analysis. Aside from the question of the simple repetition or alternation of the quantifiers as tokens, the strong parallelism of premise statements requires that account also be taken of continuation or contrast with respect to the properties they represent. This is further complicated by the peculiarities of *Some...not*, which appears in part in both subject and predicate, shares its subject token with *Some*, and expresses its scope and polarity independently. These apparent difficulties can, however, be turned to advantage and used to drive the analysis.

The nub of the problem lies in the treatment of polarity, which can be considered in either of two ways. If syntactic parallelism is the guide, then it can be thought of as a bifurcated property expressed more or less independently in both subject and predicate. Contrast with a preceding statement can then potentially occur in the polarity of either or both of these. If, on the other hand, a more traditional conception of polarity as an atomic property is adopted, then contrast is keyed to overall polarity and its locus of expression.

Using the bifurcated approach, *Some...not* is the only quantifier that is negative in the predicate, being positive and, indeed, identical to *Some* in the subject. *No*, on the other hand, is the only quantifier negative in the subject, but identical to the two positive quantifiers in the predicate. In the unitary view, they are both negative, but *Some...not* is the only quantifier that expresses its polarity in the predicate.

Tables 3.1(a) and 3.1(b) profile the 16 moods according to these two approaches. In each, the givenness or newness of the quantifier in the second premise is coded in respect of three properties. The first in both cases is the same and represents simply whether or not the quantifier in the subject is identical to the previous one. The second and third codes represent either (a) subject polarity and predicate polarity, as independent components, or (b) quantifier polarity and its locus of expression.

It must be borne in mind that the interpretation of the second- and third-place codes differs between the two tables and, whereas only the first two interact in the first table, all three interact in the second. Consequently, the absence of a star does not indicate an identical profile. What it does signify is that the practical upshot is the same, which is obvious in the case of the unitary quantifiers. Polarity expression is given in all of these instances, so any change in polarity is necessarily subsumed under a change in subject quantifier. In the remaining case of *Some...not* succeeding itself, there is of

(a) Subject identity, subject polarity, and predicate polarity

		First premise			
		All	*Some*	*No*	*Some...not*
Second premise	*All*	GGG	NGG	NNG	NGN*
	Some	NGG	GGG	NNG	GGN*
	No	NNG	NNG	GGG	NNN*
	Some...not	NGN*	GGN*	NNN*	GGG

(b) Subject identity, quantifier polarity, and locus of polarity expression

		First premise			
		All	*Some*	*No*	*Some...not*
Second premise	*All*	GGG	NGG	NNG	NNN*
	Some	NGG	GGG	NNG	GNN*
	No	NNG	NNG	GGG	NGN*
	Some...not	NNN*	GNN*	NGN*	GGG

Table 3.2: Profiles of the givenness (G) or newness (N) of the 16 moods using different conceptions of polarity. Stars indicate significant differences between the two conceptions.

course no contrast of any kind to mark.

Turning to the starred cases, the fundamental difference they highlight is the complete opposition of *Some...not* to either *All* or *No*, depending on the approach taken. When the bifurcated approach is adopted, *Some...not* proves to be the inverse of *No*, as the polarities of both their subjects and predicates differ. With the unified approach, it is the inverse of *All*, because they each express different polarities in different locations. Since this latter is arrived at using a more conventional view of polarity and its results accord with the logical semantics of the quantifiers, this is the approach that will be favoured.

3.2.2.3 Projected realisations

Combining the mood and figure profiles for a given syllogism enables its second premise to be characterised as a sequence of given and new elements. From this sequence, an 'ideal' *CC* intonation contour for the premise can be projected by employing the following assumptions, which are mainly drawn from the earlier discussion of English intonation, notably its argument-centric and phrase-final tendencies:

- The given term is deaccented (and is usually all or part of the topic).
- Newness in the quantifier or the copula leads to accentuation of those items when immediately followed by the deaccented, given term.
- Focal stress on the new term is narrow if it immediately follows a given quantifier or copula and broad otherwise; the difference may be difficult to detect in practice, but the former is apt to involve complete deaccenting of the non-term item whereas the latter is not.

Table 3.4 presents an example that illustrates the application of this method by contrasting the consequences of the two approaches to polarity previously discussed. It employs the four syllogisms that feature the transition from *No* to *Some...not* or vice versa in the two symmetrical figures. They and the intonational labelling scheme used are chosen in anticipation of the observational study reported in Chapter 4, which also features further examples using the favoured approach. The complete set of contours for all 64 syllogisms is implemented in Prolog as part of the materials for the experiments described in Chapters 6 and 7.

For each premise, the first line profiles givenness and newness, the second (italicised) gives an informal, orthographic representation in which accented words are in

	Subject		Predicate			Focus
	quantifier	end term	copula	end term		location
1.	*No*	*A*	*are*	*B*		
	Some	*C*	*are not*	*B*		
	NN	N	N	G		
	Some	**C**	are	**not**	B	
		X*?	H*		L-L%	Predicate
2.	*Some*	*A*	*are not*	*B*		
	No	*C*	*are*	*B*		
	NN	N	N	G		
	No	**C**	**are**		B	
		X*?	H*		L-L%	Predicate
3.	*No*	*B*	*are*	*A*		
	Some	*B*	*are not*	*C*		
	NN	G	N	N		
	Some	B	are not	**C**		
	X*?		*?	H*	L-L%	Predicate
4.	*Some*	*B*	*are not*	*A*		
	No	*B*	*are*	*C*		
	NN	G	N	N		
	No	B	are	**C**		
	X*?		*?	H*	L-L%	Predicate

Table 3.3: Projections of intonation patterns of second premises in four syllogisms, based on givenness and newness of elements, according to the bifurcated approach to polarity.

	Subject		Predicate			Focus
	quantifier	end term	copula	end term		location
1.	*No*	*A*	*are*	*B*		
	Some	*C*	*are not*	*B*		
	N	N	GN	G		
	Some	**C**	are **not**	B		
		X*?	H*		L-L%	Predicate
2.	*Some*	*A*	*are not*	*B*		
	No	*C*	*are*	*B*		
	NGN	N		G		
	No	**C**	are	B		
		H*			L-L%	Subject
3.	*No*	*B*	*are*	*A*		
	Some	*B*	*are not*	*C*		
	N	G	GN	N		
	Some	B	are not	**C**		
	X*?		*?	H*	L-L%	Predicate
4.	*Some*	*B*	*are not*	*A*		
	No	*B*	*are*	*C*		
	NGN	G		N		
	No	B	are	**C**		
	X*?			H*	L-L%	Predicate

Table 3.4: Projections of intonation patterns of second premises in four syllogisms, based on givenness and newness of elements, according to the unified approach to polarity.

bold, and the third is an underspecified ToBI representation of intonation.[9] *? indicates that a pitch accent may or may not be present; X*? indicates that a pitch accent is present, but does not specify which one. (For full details of ToBI labels and their use, see the ToBI conventions (Beckman and Hirschberg, 1994).)

What the example primarily shows is that the bifurcated approach is too simplistic, generating a rather clumsy stressed positive copula in *SomenotAB,NoCB*. In the unified approach, on the other hand, when *No* follows *Some...not* the *new* expression of the *given* polarity is subsumed under the *new*ness of the subject quantifier. In fact, the shift from two independently located polarities to one 'movable' polarity guarantees that the copular verb itself is never accented. In the example given, this makes the difference between subject- and predicate-focussing.

[9] Underspecification of the ToBI labelling is mainly the result of uncertainty over the intonation of marked topics, the debate over which was mentioned earlier.

Chapter 4

Prosodic analysis of spoken solutions to syllogisms

4.1 Introduction

Having derived the two classes of intonation patterns, an observational study was undertaken to determine whether or not the predicted patterns are employed by participants in the conclusion generation task and, if so, whether or not they bear any systematic relationship to the conclusions drawn. The aim was to record, transcribe, analyse, and annotate participants' spoken delivery of the premises and their conclusions as they carried out the task.

The ToBI specification (Beckman and Hirschberg, 1994) was chosen as the formalism for transcription and analysis, because, besides being well known and used, it is geared towards sentence-level pitch accents and employs a relatively small set of discrete tone labels to denote them. ToBI specifies 4 "tiers" of analysis, one each for orthography, tone labels, "break index" labels, and miscellaneous annotations. The break index tier was omitted, on the assumption that it would contribute nothing to the identification of information structure for the purposes of the study. The miscellaneous tier was used to demarcate several layers of structure, namely syllogisms, individual statements (i.e. premises and conclusions), and subjects and predicates. This facilitated automatic processing of the annotations, exploiting ToBI's decree that the nuclear pitch accent, i.e. the focal accent, is the last pitch accent in a phrase.

In spite of these advantages, the resource- and labour-intensive nature of the analysis still rendered it impractical to conduct on a large-scale. Consequently, it was decided that a small number of participants would be tested on the full set of syllo-

gisms, but only the responses to a defined subset of syllogisms would be submitted to acoustic analysis. This brought the study within the bounds of feasibility while allowing the results to be compared to previous work. The next requirement, then, was a means of defining such a subset as would reliably yield variation between participants.

Usefully detailed hypotheses were difficult to formulate at this stage, however, as the different conceptions of information packaging provide no straightforward mapping between Stenning and Cox's (2006) categories and the ones used here. Even if it were the case that credulous processors produced *CC* contours and skeptical processors produced *CN* ones, the precise relationship of these subgroups to the various combinations of RI, RO, and HOthey identified is not well defined. Given the idiosyncratic nature of the effects associated with the latter groupings, it was therefore not possible to assign clear predictions as to either contour group with any certainty. Instead, some very broad principles were drawn out regarding where in the problem set a division between *CC* and *CN* intoners might be most plainly visible.

To begin with, the analysis of given and new by figure in the last chapter showed that figure AB-BC contains a simple topic shift, such that the *CC* contour is in fact the same as the *CN*one. Moreover, the other diagonal figure, BA-CB, also generates a secondarily accented contingency reading, albeit more exaggerated. In the absence of empirical data, though, it is difficult to judge how much more exaggerated. The two symmetrical figures, on the other hand, generate markedly contrastive *CC* patterns that are clearly distinct from the *CN* one. Mood provides less obvious guidance, but it is the two negative quantifiers between them that create the greatest potential for variation.

These considerations map well onto Stenning and Cox's results. Notwithstanding their detailed findings, the influence of the diagonal figures was as strong as in every other such study. Moreover, their model also reports main effects for all the quantifiers apart from *No*, which generates differential effects among participants. Selecting only those symmetrical syllogisms which contain *No* produced a set of 14 syllogisms. Those where *No* is repeated were excluded on the grounds that its effects vary as between premises and may therefore counteract. The final set, then, contained 12 syllogisms, all of which can be seen to have generated considerable response variation in both experiments reported. They are detailed in Table 4.1.

Syllogism	1st Premise quantifier	2nd Premise quantifier	Figure	Solution quantifier	Solution term order
NoAB-AllCB	*No*	*All*	AB-CB	*No*	AC/CA
AllAB-NoCB	*All*	*No*	AB-CB	*No*	AC/CA
NoAB-SomeCB	*No*	*Some*	AB-CB	*Some...not*	CA
SomeAB-NoCB	*Some*	*No*	AB-CB	*Some...not*	AC
NoAB-SomeCnotB	*No*	*Some...not*	AB-CB	(none)	(none)
SomeAnotB-NoCB	*Some...not*	*No*	AB-CB	(none)	(none)
NoBA-AllBC	*No*	*All*	BA-BC	*Some...not*	CA
AllBA-NoBC	*All*	*No*	BA-BC	*Some...not*	AC
NoBA-SomeBC	*No*	*Some*	BA-BC	*Some...not*	CA
SomeBA-NoBC	*Some*	*No*	BA-BC	*Some...not*	AC
NoBA-SomeBnotC	*No*	*Some...not*	BA-BC	(none)	(none)
SomeBnotA-NoBC	*Some...not*	*No*	BA-BC	(none)	(none)

Table 4.1: Specifications of the 12 syllogisms used for prosodic analysis.

4.2 Method

4.2.1 Materials

The primary materials were papers setting the immediate inference and syllogistic reasoning tasks. The immediate inference paper contained all 32 possible pairings of syllogistic statements which together involve only two terms. The order of presentation of the pairings was randomised for each participant. Listed next to each pairing were the three permissible responses, TRUE, FALSE, and CAN'T TELL. Participants were instructed to indicate which of these responses applied to the second statement if the truth of the first was assumed. This was illustrated with an example in which the first statement was "No Murrels are Lellims" and the second was "Some Lellims are Murrels". They were told,

> If you decided that "Some Lellims are Murrels" must be false if "No Murrels are Lellims" is true, then you would circle *False* on the paper and say, "No Murrels are Lellims. Some Lellims are Murrels. False."

The syllogistic reasoning paper contained all 64 possible pairings of syllogistic premises which together involve three terms. The order of presentation of the pairings was randomised for each participant. Each pairing was followed by a line on which

to write a conclusion. Participants were instructed to indicate what third statement of the same form must also be true if the truth of the two premises was assumed, or else that there was no valid conclusion. This was illustrated with an example in which the premises were "All Nallims are Marrens" and "No Nallims are Lorroms". They were told,

> If you decided that, say, "Some Lorroms are Marrens" must be true if both "All Nallims are Marrens" and "No Nallims are Lorroms" are true, then you would write "Some Lorroms are Marrens" above the underline on the paper and say, "All Nallims are Marrens. No Nallims are Lorroms. Some Lorroms are Marrens".
>
> If, however, you decided that no conclusion could be drawn about Marrens and Lorroms, then you would write "No Valid Conclusion" above the underline and say, "All Nallims are Marrens. No Nallims are Lorroms. No Valid Conclusion."

Aside from fully randomising the order of the immediate inference paper, there were a number of other differences between these materials and those in Stenning and Cox (2006). First of all, the earlier experiment used single letters to denote entities in premises and conclusions, e.g. *Some As are Bs*. When spoken, though, these produce signals that are very difficult for software to calculate accurate intonation contours from. What are required are full words containing open vowels and voiced consonants. However, using actual, meaningful words would introduce the possibility of conclusions being influenced by belief bias (see Chapter 1). Consequently, nonsense words were used, e.g. *Some Mullums are Nannels*.[1]

Secondly, Stenning and Cox (2006) reused the same three letters, *A*, *B*, and *C*, in every syllogism and, moreover, assigned them the same rôles throughout: *A* was the end term in the first premise, *C* that in the second, and *B* the middle term. This confuses the standard variables used in the specification of figure with the tokens used in particular instances of syllogisms. Whether or not it had a significant impact on the results of the earlier experiment by way of a practice effect is not known, but in any case it could not be followed here, as it might interfere with participants' own assignment of discourse-old and discourse-new intonation patterns. In the procedure finally used, each term was assigned a nonsense word randomly from a pool of 53 and each word used was barred from reuse for a minimum of 4 syllogisms thereafter. For the sake of consistency and so as to accustom participants to the forms of words used, the same procedure was used to generate the immediate inference papers.

[1] I am grateful to Prof. Bob Ladd for his timely advice on this matter and his examples of appropriate nonsense words. Any deficiencies in the set of words finally used are due to me.

Lastly, in the earlier experiment, premises and conclusions were arranged vertically with respect to one another on the page and with solid horizontal lines separating the former from the latter, in the manner of mathematical problems. Since the aim of the present study was to access, if possible, participants' *linguistic* interpretations, premises and conclusions were presented sequentially on the same line, in the manner of sentences in a text.

4.2.2 Participants

Participants were 13 students at the University of Edinburgh who had responded to an advertisement placed in the University's Student Employment Service. The responses of a fourteenth participant were discarded on the grounds of failure to adhere to the rubric.

They were required to be native speakers of British English and have no formal schooling in logic. The first of these was necessitated by the differences in intonation between varieties of English. They were each paid 5 pounds for taking part.

4.2.3 Procedure

The basic form of the task was the same as that of Stenning and Cox (2006), but the spoken responses of the participants were recorded and analysed as well as their written answers. Each participant was seated alone in a sound-proofed recording studio for the duration of the task. This was to guarantee that the sound quality of the speech signal was high enough to permit analysis by software. They were presented with the immediate inference paper and given a sound check in order to set the recording level. They were then required to complete the paper in their own time, sound being recorded throughout. When they had done so, it was removed and replaced with the syllogistic reasoning paper, which they were again required to complete in their own time with sound recording throughout. Nearly all participants took approximately an hour to complete the two.

For both papers, the participant was instructed to respond to each problem first by circling or writing down their conclusion on the paper and then by saying aloud both the given statements and their conclusion. This instruction was added following trials of the materials with 4 participants, all of whom read out the given statements as soon as they were encountered in a highly repetitive manner which indicated no interpretation of them, as though"announcing" the problems before engaging with their

contents. From this it was concluded that attempting to access *processing* through intonation was doomed to failure and the study's target intonations should be those associated with the *products* of participants' reasoning, i.e. those that reflected the structures that participants had settled on in reaching their conclusions.

The written responses to both papers were marked and recorded. The spoken responses to the selected subset of 12 syllogisms were annotated according to the ToBI guidelines using the `xwaves` speech analysis software package produced by Entropic Research Laboratory Inc. (Shore, 1988; Talkin, 1989). Standard software tools[2] were then used to collate, order, and format the annotations. The processing included automatic identification and labelling of sentential focus. Samples of annotation files before and after processing are shown in the Appendix.

4.2.3.1 Analysis coding

A two-stage operationalisation of the concepts of focus and topic was used to assess the fit of the processed annotations to the theoretical projections produced in Chapter 3. Exact fits could not reasonably be relied upon, due partly to annotator error and partly to the artificiality of the task conditions resulting in unnaturalness in participants' speech. This could be manifested in forced, overstressed, or over-enunciated responses or, conversely, in flat and apparently undifferentiated deliveries. Focus was used as a first cut analysis, while topicalisation allowed a more detailed look at selective deaccenting.

Focus was assessed with reference to its location in the subject or predicate in both premises and the conclusion of each syllogism. As previously noted, ToBI deems the last pitch accent to be the focal one and the annotation files were processed accordingly. The projected second premise contours for the syllogisms in the subset were analysed on the same basis for comparison. These projected locations are shown in Tables 4.2 and 4.3.

Topicalisation was coded with reference to the accenting pattern of the constituent in the second premise that contained the middle term. The code consists of a single letter prefix denoting the constituent, followed by a two-digit number denoting the accent pattern. The prefix is T (Topic) if the relevant constituent does not contain the focus and C (Comment) otherwise. The digits are binary, denoting the presence (1) or absence (0) of an accent on the two elements in the constituent, in the order in which

[2] `sed`, `gawk`, and `sort` under UNIX

Syllogism	Subject		Predicate			Focus	Topicalisation
	quantifier	end term	copula	end term		location	code
NoAB-	NNG	N		G			
AllCB	All	**C**	are	B			
		H*			L-L%	Subject	T00
AllAB-	NNG	N		G			
NoCB	No	**C**	are	B			
		H*			L-L%	Subject	T00
NoAB-	NNG	N		G			
SomeCB	Some	**C**	are	B			
		H*			L-L%	Subject	T00
SomeAB-	NNG	N		G			
NoCB	No	**C**	are	B			
		H*			L-L%	Subject	T00
NoAB-	N	N		GN	G		
SomeCnotB	Some	**C**	are	**not**	B		
		X*?		H*	L-L%	Predicate	C10
SomeAnotB-	NGN	N		G			
NoCB	No	**C**	are	B			
		H*			L-L%	Subject	T00

Table 4.2: Projected intonation patterns of second premises in the 6 figure AB-CB syllogisms selected, based on givenness and newness of elements.

Syllogism	Subject		Predicate			Focus	Topicalisation
	quantifier	end term	copula	end term		location	code
NoBA-	NNG	G		N			
AllBC	**All**	B	are	**C**			
	X*?			H*	L-L%	Predicate	T10
AllBA-	NNG	G		N			
NoBC	**No**	B	are	**C**			
	X*?			H*	L-L%	Predicate	T10
NoBA-	NNG	G		N			
SomeBC	**Some**	B	are	**C**			
	X*?			H*	L-L%	Predicate	T10
SomeBA-	NNG	G		N			
NoBC	**No**	B	are	**C**			
	X*?			H*	L-L%	Predicate	T10
NoBA-	N	G	GN	N			
SomeBnotC	**Some**	B	are not	**C**			
	X*?		*?	H*	L-L%	Predicate	T10
SomeBnotA-	NGN	G		N			
NoBC	**No**	B	are	**C**		Predicate	T10
	X*?			H*	L-L%		

Table 4.3: Projected intonation patterns of second premises in the 6 figure BA-BC syllogisms selected, based on givenness and newness of elements.

they appear. So, an unmarked topic is coded T00, while one marked with stress on a quantifier is T10.

The codes for the projected contours are also shown in 4.2 and 4.3. It should be noted that this method identifies the focal constituent in syllogism *NoAB-SomeCnotB*, according to ToBI rules. This is acceptable, however, as the copular stress there is both salient and distinctive. The coding for the true topic would be T01. Similarly, neutral intonation in figure AB-CB would be coded C01, but in figure BA-BC it would be T00. Overall, the projected concordant and neutral codes account for only five out of a possible eight.

In the first cut of the data, the focus and constituency labelling in the processed annotation files was used to generate summaries of focus placement in the subjects or predicates of premises and conclusions. Then, the individual tone labels, together with the constituency labels, were used to classify second premises according to the topicalisation coding scheme.

4.3 Results

4.3.1 Standard measures

Table 4.4 compares the effects of structural factors on written performance of the syllogistic reasoning task in this study with representative figures taken from Johnson-Laird and Bara (1984) and Inder (1987), as summarised in Polk and Newell (1995). 1984 figures are for the untimed condition. Correctness is calculated with respect to legal responses only, i.e. excluding those containing the middle term or an invalid quantifier. Atmosphere is calculated with respect to legal, non-NVC responses, conversion with respect to incorrect, legal, non-NVC responses, and the figural effect with respect to legal, non-NVC responses in figures AB-BC and BA-CB only.

The results are broadly comparable, suggesting that the modifications to the task made for the purposes of this study did not unduly affect performance of it. Likewise, although the figures for the selected subset of 12 syllogisms are lower across the board than for the full set of 64 and especially for conversion errors, they still follow the same pattern.

Study	N	Correct %	Atmosphere %	Conversion %	Figural %
Johnson-Laird and Bara (1984)	20	40	69	28	90
Inder (1987)	3	61	84	54	70
Present (64 syllogisms)	13	61	83	43	72
Present (12 syllogisms)	13	55	77	27	n/a

Table 4.4: Comparison of structural effects in the present study with prior results

4.3.2 Focussing and topicalisation

Table 4.5 shows the breakdown of focussing behaviour in the three constituent statements of a syllogism. First premises are overwhelmingly predicate-focussed (96.8%), as are conclusions when they are drawn (96.8%). In contrast, a small but significant minority of second premises are subject-focussed (18.7%). Focussing in second premises is cross-tabulated with topicalisation behaviour in Table 4.6. There, subject-focussing is seen to be exclusively associated with code T00 and, with only one exception, the inverse is also true. Of the remaining cases, only three could not be classified under one of the eight possible codes. The four predicted codes (T00, T10, C10, C01) account for the great majority of responses (79.4%) and only one unpredicted code (T11) features significantly (14.2%).

In Table 4.7, topicalisation is broken down by syllogism. There is a virtually complete division of response types between the two figures. C10, C01, and T00 occur, with one exception, entirely in figure AB-CB, while T10 and T11 occur entirely in figure BA-BC. All four predicted codes occur in the figures where they were predicted to occur and, conversely, do not appear where they were not predicted to appear. T00,

Focus location	Statement		
	1st Premise	2nd Premise	Conclusion
Subject	5	29	3
Predicate	150	126	92
NVC response	-	-	57
No data	1	1	4
Total	156	156	156

Table 4.5: Focussing in premises and conclusions

Focus location	Topicalisation code										Total
	C00	C10	C01	C11	T??	T00	T10	T01	T11	—	
Subject						29					29
Predicate		12	32	4	2	1	49	4	22		126
No data										1	1
Total		12	32	4	2	30	49	4	22	1	156

Table 4.6: Focussing and topicalisation in second premises

(a) Figure AB-CB

Syllogism	Topicalisation code									Total
	C10	C01	C11	T??	T00	T10	T01	T11	—	
NoAB-AllCB		8			5					13
AllAB-NoCB		4	1		8					13
NoAB-SomeCB	1	6			6					13
SomeAB-NoCB		7	1		5					13
NoAB-SomeCnotB	9		2		1				1	13
SomeAnotB-NoCB	2	7			4					13
Total	12	32	4		29				1	78

(b) Figure BA-BC

Syllogism	Topicalisation code									Total
	C10	C01	C11	T??	T00	T10	T01	T11	—	
NoBA-AllBC					1	8	1	3		13
AllBA-NoBC				1		6		6		13
NoBA-SomeBC				1		9	1	2		13
SomeBA-NoBC						9		4		13
NoBA-SomeBnotC						9	1	3		13
SomeBnotA-NoBC						8	1	4		13
Total				2	1	49	4	22		78

Table 4.7: Distribution of patterns of topicalisation between syllogisms

though, which was the only one predicted to occur in both figures, appears just once in BA-BC. This confirms the implications of the relationship revealed in Table 4.6 that T00 effectively does not appear in neutral intonation patterns and subject-focussing occurs exclusively in figure AB-CB.[3] Calculated with respect to that figure alone, then, subject-focussing occurs in over a third of cases (37.7%). Given that T10 is the predicted concordant code for BA-BC and T11 was not predicted, the latter would appear to correspond to neutral intonation of that figure, overstressed.[4]

In both figures, the use of concordant intonation outweighs neutral, but less so in AB-CB than BA-BC(56.2% against 69%; 62.5% overall). There, the imbalance is fairly consistent, both in absolute and relative terms, with the concordant outnumbering neutral by roughly 2 or 3 to 1. *NoBA-SomeBC* perhaps has a particularly small number of neutral responses, but the only real exception is *AllBA-NoBC*, where the numbers are equal. In AB-CB, the imbalance can largely be accounted for by C10, but nonetheless the distribution between the other concordant pattern, T00, and the neutral, C01, is more uneven. There is arguably a general tendency in favour of C01 which is countered in *NoAB-SomeCB* and reversed in *AllAB-NoCB*. C10 was predicted for just one syllogism, *NoAB-SomeCnotB*, and its only really significant appearance is in just that one, for which it is the clearly dominant pattern.

Moving on to reasoning performance associated with intonational behaviour, Table 4.8 shows the distribution of NVC responses. NVC accounts for nearly two fifths of all responses (38.5%). The overall distribution is organised almost exactly according to mood: those syllogisms featuring *Some...not*, for which it is the correct answer, elicit the most NVC responses, followed by *Some*, while those in *All* garner the least. There is a more diffuse figural bias, with responses to figure BA-BC syllogisms ranking broadly higher up the table than those to AB-CB, reflecting the overall majority of NVC responses in that figure (58.3%).

Half of all NVC responses are delivered after one of the three major concordant

[3]Additional analysis of the patterns involved reveals that the single, exceptional case of T00 in figure BA-BCin fact conforms exactly to the paradigm of neutral intonation, with just a solitary pitch accent signalling focus on the sentence-final term. Of the 29 occurrences in figure AB-CB, only two do not conform to the projected pattern focussing the subject term, having instead narrow focus on the quantifier.

[4]Or consistently misanalysed by the annotator. The difficulty in distinguishing between broad and narrow stress combined with the likelihood of over-stressing lends considerable support to the idea that the remaining unpredicted patterns, C11 and T01, should be assimilated to C01 and T11, respectively. However, given the consistency evidenced by these results and the marginal differences involved, this approach has not been followed.

[5]This column is for missing or incomplete data, i.e. the combined figures for the columns previously labelled T?? and —.

Syllogism	Topicalisation code								Total
	C10	C01	C11	T00	T10	T01	T11	???[5]	
NoAB-SomeCnotB	8 (9)		2 (2)	0 (1)				1 (1)	11 (13)
NoBA-SomeBnotC					6 (9)	1 (1)	3 (3)		10 (13)
SomeAnotB-NoCB	1 (2)	4 (7)		3 (4)					8 (13)
SomeBnotA-NoBC					4 (8)	1 (1)	3 (4)		8 (13)
SomeBA-NoBC					4 (9)		3 (4)		7 (13)
NoBA-SomeBC					2 (9)	1 (1)	1 (2)	0 (1)	4 (13)
SomeAB-NoCB		2 (7)	0 (1)	1 (5)					3 (13)
NoBA-AllBC				0 (1)	1 (8)	1 (1)	1 (3)		3 (13)
AllBA-NoBC					1 (6)		2 (6)	0 (1)	3 (13)
NoAB-SomeCB	0 (1)	2 (6)		0 (6)					2 (13)
NoAB-AllCB		1 (8)		0 (5)					1 (13)
AllAB-NoCB		0 (4)	0 (1)	0 (8)					0 (13)
Total	9 (12)	9 (32)	2 (4)	4 (30)	18 (49)	4 (4)	13 (22)	1 (3)	60 (156)

Table 4.8: Distribution of NVC responses by syllogism and topicalisation code, in reverse order by total. (Overall totals in brackets.)

intonations of second premises (51.7%) and just over a third after one of the two neutral ones (36.7%). Proportionately, however, the latter account for more than the former (40.7% against 34.1%).

Amongst the concordant codes, only C10 shows a strong tendency towards NVC (75.0%), which is the correct answer in its limited sphere of operation. Both T00 and T10 lean in the opposite direction (13.3% and 36.7%, respectively), but with markedly different distributions. The bulk of T00 responses is gathered at the base of the table but includes no NVCs. What few of these there are are in the minority found higher up. T10 responses, by contrast, are more evenly spread down the table and display the more gradual tailing off of NVCs seen in the overall totals. That is, NVC responses following T00 topicalisation are effectively limited to those syllogisms where it is the correct answer, whereas T10 NVCs are no more discriminating in this regard than the aggregate.

Turning to the neutral codes, C01 can be characterised in much the same terms as T10 is above. Although in total it contributes a smaller proportion of NVC responses (28.1%), its general pattern of distribution is very similar in size and shape, differing only in a slightly more rapid tailing off. T11 also shares the shape, but this time in

Syllogism ID	Topicalisation code								Total
	C10	C01	C11	T00	T10	T01	T11	???	
NoAB-AllCB		6 (8)		5 (5)					11 (13)
AllAB-NoCB		2 (4)	1 (1)	8 (8)					11 (13)
NoAB-SomeCnotB	8 (9)		2 (2)	0 (1)				1 (1)	10 (13)
NoBA-SomeBnotC					6 (9)	1 (1)	3 (3)		10 (13)
SomeAnotB-NoCB	1 (2)	4 (7)		3 (4)					8 (13)
SomeBnotA-NoBC					4 (8)	1 (1)	3 (4)		8 (13)
NoAB-SomeCB	1 (1)	3 (6)		3 (6)					7 (13)
NoBA-SomeBC					4 (9)	0 (1)	0 (2)	0 (1)	4 (13)
AllBA-NoBC					3 (6)		0 (6)	1 (1)	4 (13)
SomeBA-NoBC					4 (9)		0 (4)		4 (13)
SomeAB-NoCB		1 (7)	0 (1)	2 (5)					3 (13)
NoBA-AllBC				1 (1)	0 (8)	0 (1)	1 (3)		2 (13)
Total	10 (12)	16 (32)	3 (4)	22 (30)	21 (49)	2 (4)	7 (22)	2 (3)	83 (156)

Table 4.9: Distribution of correct responses by syllogism and topicalisation code, in reverse order by total. (Overall totals in brackets.)

the context of much higher proportions of NVC responses throughout (59.1% overall). Bearing in mind there is a general figural bias, it may therefore be noted that, within their respective figures, neutral codes are associated with much higher proportions of incorrect NVC responses than concordant ones.

Following on from these observations, then, Table 4.9 shows the distribution of correct responses between codes and syllogisms. Just over half of all responses are correct (53.2%) and over three fifths of these relate to syllogisms in figure AB-CB (62.7%). However, the results are again organised largely by mood. The table is all but bracketed by syllogisms in *All*, with those in figure AB-CB at the top and those in BA-BC at the bottom. Inside the brackets, syllogisms in *Some...not* rank above those in *Some*.

The three concordant topicalisation codes account for nearly two thirds of all correct responses (63.9%) while the two neutrals take up over a quarter (27.7%). Proportionately, the success rate of the former is also much higher than the latter (58.2% against 42.6%).

The highest scoring concordant code is C10 (83.3%), due entirely to its concentration of NVC responses where this is correct, observed above. T00 also fares very

well (73.3%), this time due primarily to the two AB-CB syllogisms in *All*. The two in *Some* garner almost as many responses but only half of them are correct, while those in *Some...not* are significantly underrepresented but, on the whole, high-scoring. T10 is the least successful of the three, being the only one to achieve less than half marks (42.9%). This, though, is effectively due entirely to syllogism *NoBA-AllBC*, at the bottom of the table; the rest of the marks are distributed rather evenly at a ratio very close to 2:1.

C01 is the higher scoring of the two neutral patterns, even though only half of the conclusions following it are correct (50.0%). This ratio persists into the detailed figures, except for the highest and lowest, which counterbalance. With only a third of all responses correct (31.8%), T11 scores the lowest of all five major patterns. Interestingly, although its responses are fairly well spread, the correct ones occur almost entirely in the two syllogisms in *Some...not*. Overall, then, and especially considering the figural bias, it may be noted that concordant patterns are associated with much greater proportions of correct answers than neutral ones.

The gist of the above findings is summarised in Table 4.10. This clearly illustrates the complementarity of C10 and T00, with the former strongly associated with correct NVC responses and the latter with correct non-NVC responses. C01, in contrast, shows no comparable association with response type or correctness. The greater difficulty of BA-BC syllogisms is apparent in the scores for T10 and T11, but the former greatly exceeds the latter, particularly in the number and accuracy of its non-NVC responses.

Turning now to conclusion form, Table 4.11 shows the distribution of conclusions whose terms appear in the order AC compared to those in either of the two valid term orders. In total, exactly half are so ordered (50.0%), with only slightly more of these coming from the figure BA-BC syllogisms than from AB-CB (53.3% against 46.7%). Aside from this minor figural imbalance, however, there is little obvious organisation

Response	Topicalisation code								Total
Type	C10	C01	C11	T00	T10	T01	T11	???	
NVC	9 (9)	4 (9)	2 (2)	3 (4)	10 (18)	2 (4)	6 (13)	1 (1)	37 (60)
Non-NVC	1 (3)	12 (23)	1 (2)	19 (26)	11 (31)		1 (9)	1 (2)	46 (96)
Total	10 (12)	16 (32)	3 (4)	22 (30)	21 (49)	2 (4)	7 (22)	2 (3)	83 (156)

Table 4.10: Summary of correct NVC and non-NVC responses by topicalisation code. (Overall totals in brackets.)

Syllogism ID	Topicalisation code C10	C01	C11	T00	T10	T01	T11	???	Total
AllBA-NoBC					5 (5)		3 (4)	1 (1)	9 (10)
SomeAnotB-NoCB		2 (3)		1 (1)					3 (4)
SomeBA-NoBC					4 (5)		0 (1)		4 (6)
NoBA-SomeBC					3 (7)		1 (1)	1 (1)	5 (9)
NoBA-AllBC				0 (1)	4 (7)		1 (2)		5 (10)
AllAB-NoCB		0 (4)	1 (1)	5 (8)					6 (13)
SomeAB-NoCB		1 (3)	0 (1)	2 (3)					3 (7)
SomeBnotA-NoBC					1 (4)		1 (1)		2 (5)
NoAB-AllCB		2 (6)		2 (5)					4 (11)
NoBA-SomeBnotC					1 (3)				1 (3)
NoAB-SomeCB	0 (1)	1 (4)		2 (5)					3 (10)
NoAB-SomeCnotB	0 (1)			0 (1)					0 (2)
Total	0 (2)	6 (20)	1 (2)	12 (24)	18 (31)		6 (9)	2 (2)	45 (90)

Table 4.11: Distribution of AC conclusions by syllogism and topicalisation code, in reverse ratio order by total. (Overall totals in either valid term ordering in brackets.)

in the overall distribution. However, a considerable disparity is revealed when the neutral and concordant patterns are compared. The latter provide nearly two thirds of the well-formed responses (63.3%) whereas the former supply only one third (32.2%). There is less between them, though, when their respective proportions of AC responses are considered (52.6% and 41.4%).

Looking at the concordants individually, C10 is negligible. The T00 responses are split evenly according to whether *No* is in the first premise or the second. The overwhelming majority of the former are themselves evenly divided between the syllogisms with *All* and *Some* in the other premise and show a tendency away from AC ordering (40.0%), whereas almost all of the latter occur with *All* and most are AC (63.6%). Something similar can be said of T10, except that the basic split is weighted slightly towards *No* in the first premise, where there is no preference for term ordering (50.0% AC), while *No* in the second premise elicits a strong tendency towards AC conclusions (90.0%), and there is a small but significant set of responses linked to *Some...not* in either premise which manifests a strong tendency against AC ordering (28.6%).

Of the neutral codes, C01 displays an overall inclination away from AC conclusions (30.0%). This is reflected in the detail, where the responses are distributed more or less

evenly. The only exception is in syllogism *SomeAnotB-NoCB*, where it is effectively reversed (66.7%). T11, in contrast, shows exactly this preference overall (66.7%). Half of this is attributable to just the one syllogism, *AllBA-NoBC*, with the other half accrued sparely but evenly across the remainder. In fact, T11 supplies exactly the same number of AC responses as C01 but only half the number of well-formed conclusions overall.

Finally, it remains to consider the behaviour of participants as individuals. Table 4.12 shows the numbers of topicalisation codes broken down this time by participant, along with their scores for rash and hesitant behaviour in the immediate inference task. Surprisingly, there is no strong distinction between participants in terms of their intonational preferences, use of all five significant patterns being spread broadly across the sample. No participant uses fewer than three of them and all bar two favour concordant over neutral patterns to greater or lesser degrees. None of the interpretational traits from Stenning and Cox (2006) appears to have any bearing. For the sake of completeness, Tables 4.13, 4.14, and 4.15 give the totals by participant of NVC, correct, and AC conclusions, respectively, similar to those given above for individual syllogisms. They likewise reveal no clear differentiation in the participant population.

Participant	Hesitant/Rash			Topicalisation code								
	HO	RI	RO	C10	C01	C11	T00	T10	T01	T11	???	Total
1	2	2	1		4	2		1		5		12
2	1	4	5	1	1		4	2	2	2		12
3	0	4	7	1	1		4	2		3	1	12
4	1	4	4	2	2		2	5		1		12
5	1	2	0	2			4	5		1		12
6	0	3	4	1	4		1	3	2		1	12
7	0	0	5	1	1		5	5				12
8	2	1	2		3	1	2	4		2		12
9	0	4	8	1	4		1	5		1		12
10	0	0	1	1	2		3	5		1		12
11	3	4	2	1	3	1	1	5		1		12
12	0	2	4		5			3		3	1	12
13	5	4	4	1	2		3	4		2		12
Total	-	-	-	12	32	4	30	49	4	22	3	156

Table 4.12: Distribution of patterns of topicalisation between participants.

Participant	Topicalisation code								
	C10	C01	C11	T00	T10	T01	T11	???	Total
4	2 (2)	2 (2)		0 (2)	5 (5)		1 (1)		10 (12)
1		1 (4)	1 (2)		1 (1)		5 (5)		8 (12)
6	1 (1)	3 (4)		0 (1)	2 (3)	2 (2)		0 (1)	8 (12)
8		1 (3)	1 (1)	0 (2)	2 (4)		2 (2)		6 (12)
3	1 (1)	0 (1)		2 (4)	0 (2)		2 (3)	0 (1)	5 (12)
12		1 (5)			1 (3)		2 (3)	1 (1)	5 (12)
13	1 (1)	0 (2)		1 (3)	2 (4)		1 (2)		5 (12)
7	1 (1)	0 (1)		1 (5)	2 (5)				4 (12)
2	0 (1)	0 (1)		0 (4)	1 (2)	2 (2)	0 (2)		3 (12)
11	1 (1)	1 (3)	0 (1)	0 (1)	1 (5)		0 (1)		3 (12)
10	1 (1)	0 (2)		0 (3)	1 (5)		0 (1)		2 (12)
9	1 (1)	0 (4)		0 (1)	0 (5)		0 (1)		1 (12)
5	0 (2)			0 (4)	0 (5)		0 (1)		0 (12)
Total	9 (12)	9 (32)	2 (4)	4 (30)	18 (49)	4 (4)	13 (22)	1 (3)	60 (156)

Table 4.13: Distribution of NVC responses by participant and topicalisation code, in reverse order by total. (Overall totals from Table 4.12 in brackets.)

Participant	Topicalisation code								
	C10	C01	C11	T00	T10	T01	T11	???	Total
7	1 (1)	1 (1)		4 (5)	5 (5)				11 (12)
12		3 (5)			2 (3)		3 (3)	1 (1)	9 (12)
1		4 (4)	2 (2)		0 (1)		2 (5)		8 (12)
10	1 (1)	1 (2)		2 (3)	4 (5)		0 (1)		8 (12)
2	0 (1)	1 (1)		3 (4)	1 (2)	2 (2)	0 (2)		7 (12)
3	1 (1)	1 (1)		3 (4)	0 (2)		2 (3)	0 (1)	7 (12)
8		2 (3)	1 (1)	1 (2)	3 (4)		0 (2)		7 (12)
4	2 (2)	0 (2)		2 (2)	2 (5)		0 (1)		6 (12)
6	1 (1)	2 (4)		1 (1)	1 (3)	0 (2)		1 (1)	6 (12)
13	1 (1)	0 (2)		2 (3)	2 (4)		0 (2)		5 (12)
5	1 (2)			2 (4)	1 (5)		0 (1)		4 (12)
9	1 (1)	1 (4)		1 (1)	0 (5)		0 (1)		3 (12)
11	1 (1)	0 (3)	0 (1)	1 (1)	0 (5)		0 (1)		2 (12)
Total	10 (12)	16 (32)	3 (4)	22 (30)	21 (49)	2 (4)	7 (22)	2 (3)	83 (156)

Table 4.14: Distribution of correct responses by participant and topicalisation code, in reverse order by total. (Overall totals from Table 4.12 in brackets.)

Participant	Topicalisation code								
	C10	C01	C11	T00	T10	T01	T11	???	Total
13				2 (2)	2 (2)		1 (1)		5 (5)
1		2 (3)	1 (1)						3 (4)
6		0 (1)		1 (1)	1 (1)			1 (1)	3 (4)
3		0 (1)		2 (2)	2 (2)		0 (1)	1 (1)	5 (7)
10		1 (2)		1 (2)	3 (4)		1 (1)		6 (9)
2	0 (1)	0 (1)		2 (4)	1 (1)		2 (2)		5 (9)
4				1 (2)					1 (2)
7		0 (1)		2 (4)	2 (3)				4 (8)
9		2 (4)		0 (1)	2 (5)		1 (1)		5 (11)
5	0 (1)			1 (4)	2 (5)		1 (1)		4 (11)
8		0 (1)		0 (1)	1 (2)				1 (4)
11		1 (2)	0 (1)	0 (1)	1 (4)		0 (1)		2 (9)
12		0 (4)			1 (2)		0 (1)		1 (7)
Total	0 (2)	6 (20)	1 (2)	12 (24)	18 (31)		6 (9)	2 (2)	45 (90)

Table 4.15: Distribution of AC conclusions by participant and topicalisation code, in reverse ratio order by total. (Overall totals in either valid term ordering in brackets.)

4.4 Discussion

The results support well the general hypothesis that the assignment of information structure, as revealed by intonation, is a factor in the performance of syllogistic reasoning by untrained participants. That is, there is systematic variation in participants' patterns of intonation when reciting syllogisms they have completed and this is reflected in differences in their answers to them. As expected, variation in sentential focussing occurs in second premises but not first and in figure AB-CB but not BA-BC, while concomitant variation in the intonation of topics occurs in the second premises of both figures and is linked in surprisingly detailed fashion to structural features that create distinctive patterns of continuation from and contrast with first premises. Also, the use of *CC* intonation is in general associated with a greater degree of conclusion drawing than the use of *CN* and, in fact, turns out to be associated with a greater degree of correctness as well. The one major surprise is that information structure sensitivity, as evidenced by intonational variation and its associated effects, is a trait that differentiates between individuals in the participant population.

In spite of this last finding, the behaviours uncovered in this study are distinct in character. C10 clearly represents a very specific response to the negated predicate which leads to the correct answer that no conclusion follows. T00 is the complementary response, by and large, which registers the linking term when a conclusion does follow and typically produces the right one. C01, in contrast, shows complete insensitivity to informational status and relative lack of discrimination in the answers it generates. Indeed, although its performance is still significantly better than chance, compared to C10 and T00 it seems almost like guesswork. T10 might be described similarly, except that it is on a generally harder set of problems that it equals the scores of C01 and it does so with acknowledgement of the status of the linking term. T11, on the other hand, appears to carry the hallmarks of a general "don't know" response in its disregard of information status, much poorer performance, and extensive reliance on NVC. Whereas with C10 and T00, at least, NVC seems to be employed positively as the right conclusion, with T11 it suggests a failure to conclude which is sometimes merely accidentally correct.

In short, these responses might even be paraphrased as "There is no conclusion." (C10), "There is a conclusion and it is ..." (T00), "I think the conclusion might be ..." (T10), "I don't really know but I'm going to say the conclusion is ..." (C01), and "I don't know what the conclusion is." (T11). It would be very easy to conceive of such

behaviours being exhibited by groups of participants with different tendencies, but the results do not bear this out. It cannot be said that some of these participants respect latent information structure more than others, nor that some find different structures from others. Rather, it seems that all of them express the appropriate structure as and when they find such coherence between the premises as generates an answer of one sort or another. This points instead to the various intonational behaviours perhaps being common indicators of at least some of the different outcomes of reasoning at which any participant might arrive, instead of representing triggers that produce those outcomes.

There is then something of a puzzle here. Neither Stenning and Cox's (2006) conception of information packaging nor the one employed in this thesis appear to provide a fully adequate, or an adequately full, account. While the latter keys into the substance of conclusion drawing but fails to demonstrate causation, the former evinces a positive effect but only of one part of information packaging and largely confined to the form of the conclusion drawn. Either the two are not truly addressing the same issue, or there is a piece of the puzzle remaining to be found.

The first of these possibilities requires Stenning and Cox's notion of grammar to be distinguished more clearly from information structure as being purely a matter of surface ordering rather than relations between constituents or entities. Then, quite simply, ordering in premises influences ordering in conclusions, but differently according to participants' varying sensitivity to such ordering. This leaves the content of their reasoning open to expression through information structure. With regard to this, it is perhaps suggestive that such term ordering variation as is discernible in the results above is more suggestive of Stenning and Yule's (1997) Source Founding Model than Stenning and Cox's individual differences account.

Briefly put, that model's heuristics for identifying a unique source premise select, first of all, one with an existential quantifier over one with a universal quantifier and then, failing that, a positive over a negative. Applied to the subset of syllogisms used in the present study, this resolves into always choosing the premise that does not contain *No*. This results in an equal division of first and second source premises which is orthogonal to the figural distinction. Now, interestingly, it is only in relation to term ordering that premise order appears as a significant organising factor in the results above. Both of the main concordant codes are, with minor exceptions, associated with higher numbers of AC conclusions than CA conclusions where *No* is in the second premise and equal or lower numbers where it is in the first, essentially in line with

the Source Founding Hypothesis. In contrast, the two neutral codes generate no such split in the data, instead favouring CA in figure AB-CB and AC in BA-BC more or less across the board. This may be considered a recency effect, as it places C in the same position in the conclusion as it occupied in the second premise, both grammatically and informationally.[6]

It is only fair, however, to address the alternative possibility mentioned. One advantage of Stenning and Cox's approach is that it is based on a pre-test that gives good reason to believe that the traits it identifies are indeed operative from the outset. In contrast, it was admitted above that the data in the present study represent the products of the task, not the processes involved in it. This leaves it open for the missing piece of the puzzle to be some prior, as yet unknown factor that induces a participant to adopt a particular information structure and, as a result of that, arrive at a particular conclusion. If it were possible to pre-define the information structures in the materials, this theory could be tested. The development of the means for doing this is the subject of the next chapter.

4.5 Coda

Given how closely the patterns found in this study match the ones derived theoretically in Chapter 3 from the latent information structures of the set of syllogisms, consideration must be given to the possibility that they were found precisely because they were the ones the investigator expected or hoped to find. One drawback of the ToBI approach, is that its apparent simplicity is bought at the expense of a degree of interpretation on the part of the annotator. Given that the investigator here had no prior expertise in the art and that even those experienced in it do not necessarily produce consistent results (Steedman, 2004), this is a possibility not to be overlooked.

In point of fact, a post-test was carried out to verify the consistency and replicability of the annotations produced here, but there is insufficient space to report it in full. It comprised two parts, both utilising a subset of syllogisms drawn half from the subset of 12 analysed above and half not previously annotated. The investigator's performance proved to be consistent with his own earlier performance and that of one out of two independent annotators.

[6] It will be recalled that virtually all non-NVC conclusions are predicate-focussed.

Chapter 5

Controlling syllogism intonation using voice synthesis

5.1 Introduction

The study reported in Chapter 4 shows a connection between manifestations of information structure and performance in the syllogistic reasoning task. It is possible, however, that the intonation structures observed there are merely by-products of the reasoning process rather than integral to it. This may have been due to participants having been left free to select appropriate information structures for themselves. In order to pursue the connection further, then, it is desirable to be able to specify particular structures in the delivery of the materials. This requires the delivery to be spoken in a precisely controlled manner. There are effectively two ways in which this might be achieved. One is to employ a professional speaker to read from a suitably annotated script. The other is to use a voice synthesiser whose control parameters include specification of intonation.

The logistics of employing a professional speaker, as well as the expense, are quite prohibitive. First, a number of scripts must be prepared, in order to distribute the conditions properly. Each of them would then have to be read by the speaker either to a group of participants in a venue of appropriate size and acoustics or to individual participants on a one-to-one basis. In any event, an error-free performance on the day can not be guaranteed, leading to the likelihood of inaccurate and inconsistent presentation of the stimuli across the participant population. This is all the more likely in the present context, as professional talents lie mainly in the realm of situated natural speech and its felicitous realisation, not the artificial and largely context-free utterances

involved here and their sometimes deliberately infelicitous realisation.

The cost of developing a limited domain synthetic voice, in contrast, is quite modest. Such voices can be developed rapidly, as only a restricted set of speech sounds and contexts is drawn upon. The restricted scope of the target domain also increases the quality of the output, as sounds are only used in the contexts to which they are appropriate.[1] Voice quality is an important factor in this project, in spite of the artificial nature of the utterances, as it is the influence of human natural language mechanisms that is at issue. It is for this reason that the particular synthesis method preferred is 'unit selection' (Black and Taylor, 1997). Unit selection uses a database of recorded speech to supply the building blocks for synthesised utterances. One advantage of this is that many of the features of the speech output, such as duration, pitch, and stress, do not need to be predicted or calculated, as they simply inhere in the database, in all their ill-understood subtlety. Furthermore, although it requires a human speaker to be recorded in the first place, any errors can be gone over and corrected before the voice is built. As a result, the method produces high quality, naturalistic output.

In general, though, unit selection does not allow for intonation to be specified in the input. Instead, the output carries whatever intonation is associated with the best fitting speech units that make up the utterance requested. That is, a standard unit selection synthesiser will always generate a given utterance with the same, 'optimal' intonation, irrespective of its degree of appropriateness in a given discourse context. However, FESTIVAL 2, the latest version of the FESTIVAL open source speech synthesiser (Taylor *et al.*, 1998; Clark *et al.*, 2004), includes both a general purpose unit selection engine, called *multisyn*, and support for the APML markup language. APML is an XML-based markup language for specifying the turn-taking, performative, affective, and, crucially for present purposes, intonational aspects of texts, aimed at eliciting believable spoken and facial-gestural behaviour from virtual characters (de Carolis *et al.*, 2004). Coming from the same tradition as ToBI, it enables annotation at exactly the right level of granularity for present purposes.[2]

Prior to the present work, voices exploiting these facilities had already been built to tailor the intonation of utterances to defined contexts as components of the FLIGHTS and COMIC projects (Baker *et al.*, 2004; Foster, 2004, ,respectively). Both the speaker used and the linguistic resources developed in the latter were available to the author.

[1] Of course, the drawback is that use outwith the target domain is likely to produce very poor results, but this is not a concern here.

[2] In fact, FESTIVAL includes support for ToBI labelling also. However, it is geared towards research into improving the ToBI specification rather than real world synthesis.

Furthermore, among the design criteria for multisyn are:

> "[T]he system is designed to be robust enough to be used as a real world synthesiser, rather than just a research tool which works for a few restricted examples."

and

> "[T]he voice building process is designed to be simple enough that only very limited specialist knowledge is required to build new voices."
>
> (Clark *et al.*, 2004)

The first of these was considered a desirable feature in allowing the possibility of constructing self-directed, self-paced experiments. The second was considered desirable in relation to the voice being developed by the present author, as was the local presence at the University of Edinburgh of the FESTIVAL development team. For these reasons, it was decided to use FESTIVAL 2 to construct a limited domain unit selection voice for the controlled intonation of syllogistic premises.

5.2 The voice building process

5.2.1 Overview

The phase crucial to the success of systems using unit selection is the preliminary one of constructing the database of recorded speech. Particularly important are the selection of the speaker, the specification of the set of utterances to record (the 'script'), and the annotation of the recorded utterances for linguistic structure. The full process of building a new multisyn voice for FESTIVAL 2 is a semi-automated procedure comprising the following steps:

- Choosing the speaker.
- Designing the script.
- Recording the voice.
- Specifying linguistic resources.
- Labelling the recorded speech sounds.
- Building the database of annotated utterances.

- Specifying synthesis parameters.

Each of these steps in the construction of the syllogism voice is described in the subsections that follow. The technical detail, although considerable, is beyond the scope of this work and is mostly omitted. Some points, however, are relevant to the investigation being pursued and are therefore covered in some depth, as are those that proved problematic and required additional effort to correct or modify the working of the standard system. In particular, development and testing of the voice revealed a less than complete dovetailing of the multisyn and APML modules which had to be overcome. To assist in these more detailed areas, the following paragraph briefly introduces some terminology and outlines the synthesis process.

With a suitable voice built and loaded and an output utterance requested, FESTIVAL 2 first produces a target list of the required fundamental speech sounds ('phones') annotated with the target linguistic structure. Next, pairs of adjacent phones ('diphones') from the database are formed into candidate lists for all the target diphones. Every candidate diphone is then costed for how well it can be spliced with every candidate for the preceding overlapping diphone and how well it matches the linguistic features of the target diphone. The candidate lists are then searched to find an optimal sequence that minimises the join and target costs. This favours chains of diphones from the same source utterance, for which the join cost is zero, as long as they are consistent with the target context. Next to the quality of the database, the join and target cost functions are the most important factors determining the quality of the output.

5.2.2 Choice of speaker

As has already been indicated, this choice was effectively made at the outset. For the record, however, it should be noted that the chosen speaker possessed the following essential qualities:

- a clear voice.
- a natural-sounding, intonationally-varied delivery.
- a consistent delivery.
- all of the above over periods of up to three hours.

In addition to these, the speaker brought experience of the general procedure and a degree of linguistic and domain knowledge which proved useful in understanding and performing his task.

5.2.3 Script design

The fundamental requirement of the recording script is that it provide complete coverage of the target utterance space. That is, minimally, it should feature at least one instance of each potential output diphone (Clark *et al.*, 2004).[3] Furthermore, diphone here means diphone *in context*: it is a complex combining the diphone itself with its positions in the syllable, word, and phrase in which it appears, the stressing of the syllable, the accenting and grammatical category of the word, and other relevant linguistic features, including aspects of the preceding and following contexts. Limited domain synthesis provides an advantage here, in that coverage of each *word* in context effectively ensures coverage of each diphone. The task, therefore, is to distribute each word in the domain vocabulary systematically across all its potential contexts of occurrence.

Ideally, in fact, the system should have several instances of each such occurrence to choose from, so as to enable it to maximise the join quality. However, this desideratum takes second place to that of producing a script of a manageable size and in the majority of cases this results in attempting to achieve complete coverage as parsimoniously as possible. As points of reference, the FLIGHTS script contained 149 sentences and the COMIC script 308 (Baker, 2003; Foster, 2004). Reporting on the development of the FLIGHTS system, Baker (2003) analyses the script design task into three components, each of which will be dealt with in turn in the next three subsections: variables, sentence types, and prosodic context.

Having enumerated each of these components, Baker goes on to present an algorithm for combining them in the fewest number of sentences. This is useful where, as there, there are sentence types and lists of variables of varying lengths. It is not necessary in the present case, however, as the sentence types are highly uniform and the variables completely interchangeable, making the combinatrics involved much more straightforward. The final subsection, therefore, simply deals with the manner in which the script sentences are presented to the speaker.

[3] Although FESTIVAL 2 implements a 'back-off' procedure for finding replacements for missing diphones, this necessarily leads to uncontrolled results and is to be avoided where at all possible.

5.2.3.1 Variables

These are the words, usually but not necessarily nouns or noun phrases, that instantiate situational parameters. In the FLIGHTS domain of airline bookings, they include the names of specific airlines, airports, and cities, plus times of day and so on. In the syllogism domain, they are the terms of the syllogism. Now, in the observational study, the terms were nonsense words, but that was only to assist the speech analysis software in determining intonation contours. Human listeners have much greater facility in this regard, so it was decided that known words denoting real-life categories would be used. Of course, this raises the spectre of belief bias (see Chapter 1), but the plethora of previous studies of syllogistic reasoning provide a wealth of choice of sets of terms that have been compiled with this in mind.

The set of terms chosen was that published as Table 17 in Stenning and Yule (1997). This lists three classes of terms, denoting professions, nationalities, and interests or pastimes, each containing 32 entries. As such, it is far too large to use in its entirety. Pre-empting a little the discussions to come, if it is assumed that

(a) each quantifier takes a different *CC* contour in the second premise of each figure, plus one *CN* contour for all of them, and

(b) each script premise accounts both for an instance of one term in subject position and an instance of another in predicate position and therefore terms only count once, not twice,

then there are a minimum of $4 \times 5 \times 3 \times 32 = 1920$ sentences to be recorded, without even accounting for first premises or repetition of quantifiers!

Clearly, reducing the pool of terms on which to draw increases the potential for participants becoming confused by repetition of terms in successive syllogisms, so a balance must be struck. It was decided that 8 terms from each class was the minimum acceptable number, producing 480 combinations on the assumptions above. This is still considerably more than was required for either of the earlier voices, and yet to be increased by the additional factors mentioned, but the advice received was that a script double that size could be recorded in under 3 hours, so it was deemed acceptable.

The next step was to identify which members of each class to keep. Here, it was possible to exploit what Baker (2003) calls ‘systematic overlap’, which occurs when every member of a class of words contains the same phone at the same boundary, i.e. start or end, as a regular feature and is therefore interchangeable at that boundary with

Professions	Nationalities	Interests
Butchers	Koreans	Chessplayers
Musicians	Egyptians	Beerdrinkers
Waiters	Hungarians	Hillwalkers
Judges	Canadians	Squashplayers
Nurses	Indonesians	Horseriders
Sailors	Brazilians	Golfers
Teachers	Tibetans	Tennisplayers
Busdrivers	Mexicans	Dogowners

Table 5.1: Final lists of terms included in voice, by category

every other member of the class without affecting the boundary diphone. This results in only one instance of a given context being needed in order to supply that context to all members of the class. In the syllogism case, this ensures that each term only needs to be recorded once in the subject for each context and once in the predicate, as the two can be completely cross-matched at the diphone boundary of the subject term and the copula.[4]

Since the Table 17 terms are all plural, this favours selecting those with the regular plural ending '-s'. This can be realised by either of the phones [z] or [s] according to whether the previous phone is voiced or unvoiced, respectively. Tone is carried by voice, so the [z] instances were chosen. This, however, still left 20 professions, 24 nationalities, and 26 interests. Extending the approach to '-ns' for nationalities and '-ers' for interests only reduced these to 17 and 21 respectively, so thereafter an ad hoc approach was adopted, eliminating any terms that might yet be unduly semantically loaded or over-conceptualised (e.g. "politicians", "Australians") or trigger stereotypical conceptions when in combination with certain other terms (e.g. "Russians" with "chessplayers"). The final set is shown in Table 5.1.

5.2.3.2 Sentence types

These are the sentence 'templates' that contain the slots into which variables fit. In the syllogism domain, these would seem to be clearly generated by the various combinations of the quantifiers, the copula, and the negative particle. However, the asymmetry

[4] It is, in fact, for this reason that the initial calculation given was not $4 \times 5 \times 3 \times 32 \times (3 \times 32 - 1) = 182400$!

of the quantifiers raises the question whether or not there are in fact only two sentence types - with or without the negative particle - and the three subject quantifiers are better considered as variables. In the result, the permutations are the same whether the subject quantifiers are treated as sentence types or variables, while the restriction of a negative sentence type to *Some* alone only serves to obscure the potential cross-matches between the three positively predicated quantifiers. Consequently, it was concluded that the four quantifiers generate one sentence type each.

5.2.3.3 Phonetic coverage

Having already effectively dealt with the issues relating to diphones in combination, this component of the task resolves into enumerating the set of possible intonation contours that each sentence type may carry. This, of course, was already covered in Chapter 3. Allowing for the redundancy found therein, however, eliminates 28 of the 64 contours produced there. For each of the three unitary quantifiers, there remain the same two subsets of four contours, two per figure, corresponding to whether or not the quantifier in question is the same as that in the first premise. For *Some...not*, there remain three subsets, one each for following itself, following *Some*, and following *All* or *No*.

The *CN* pattern for second premises is characterised by secondary accenting of the subject term and focussing of the predicate. This is the same as one or other of the *CC* patterns produced for each quantifier in figure AB-BC and is therefore already accounted for. First premises, being discourse-initial, are considered also to be *CN* but nonetheless generate distinct contours on account of 'continuation rise' at the end. However, being first and *CN* they are necessarily unaffected by mood or figure and therefore contribute only one further contour to the inventory of each sentence type.[5] There are, then, in total, $(3\times2\times4)+(3\times4)+4=40$ phonetic contexts to incorporate into the recording script.

Factoring in the variables, this results in $40\times24=960$ sentences in all, which, as it happens, is exactly the 2×480 contemplated earlier. That is not the end of the matter, though, because the set of phonetic contexts identified still contains redundancies that enable further cross-matching to be done. Earlier, only cross-matching within each context was contemplated, but now cross-matching between contexts proves to be possible. To illustrate this, consider the following examples:

[5]The possibility of anticipatory contrast was touched on in Chapter 3 and will be passed over here.

(5.1) a. *All* ***butchers*** *are* ***Hungarians.***
H* H* LH%

b. *All* ***Koreans*** *are* ***beerdrinkers.***
H* H* LH%

(5.2) a. *No* ***nurses*** *are squashplayers.*
H* LL%

b. *No* ***Tibetans*** *are waiters.*
H* LL%

Now, cross-matching within the contexts represented by (5.1) and (5.2) generates the instances in (5.3), among others, but cross-matching between them additionally generates the likes of those in (5.4):

(5.3) a. *All* ***butchers*** *are* ***beerdrinkers.***
H* H* LH%

b. *No* ***Tibetans*** *are squashplayers.*
H* LL%

(5.4) a. *All* ***Koreans*** *are squashplayers.*
H* LL%

b. *No* ***nurses*** *are* ***Hungarians.***
H* H* LH%

The intonational isomorphism of the contexts identified for the unitary quantifiers and the obvious overlap between *Some* and *Some...not*enable a considerable amount of such cross-matching to be done. The constraining factor is ensuring coverage of all the variables in the subject phrase of each context, while the greatest slack is found in the contexts for *Some*, which overlap those for the other two unitary quantifiers as well as those for *Some...not*. In the result, one each of the contexts for *All* and *No*, four of those for *Some...not*, and five of those for *Some* could be dispensed with. The final tally of sentences to be included in the recording script therefore comes to $29 \times 24 = 696$.

5.2.3.4 Script presentation

The script was generated automatically using a Prolog program to combine the sets of variables with the set of phonetic contexts. It was printed for presentation to the

speaker, with each page containing half of the 24 sentences in a particular context. This was to enable the speaker to familiarise himself with each context in turn and record all the sentences in it together, for maximum consistency.

The intonational contour of each sentence was represented by typographical cues. Experience from the COMIC project had shown that this was easier for the speaker to interpret than annotating the words with analytical symbols. Bold face was used for standard pitch accents, while italics indicated exaggerated emphasis. Punctuation was used to signal boundary tones, with full stops indicating low tones and ellipsis signifying continuation rise. Examples 5.5 and 5.6 illustrate these devices for a first and a second premise context, respectively.

(5.5) **All Egyptians** are **musicians** ...

(5.6) Some *teachers* are not **dogowners**.

As indicated in Chapter 3, many if not all of the script sentences are likely, strictly speaking, to be composed of more than one intermediate phrase, each with its own boundary tone. Given their brevity and structural simplicity, however, pitch accents and final boundary tones were considered sufficient to convey the intended contours to the speaker and intermediate tones, indeed, feared intrusive and confusing. In effect, the speaker's own linguistic intuitions were trusted to find the most natural phrasing in each case.

In the FLIGHTS script, each sentence was prefaced by a context-setting question, delivered during recording by another interlocutor. This suited the overall conversational dialogue context for which the voice was intended. It was not considered helpful in the present case, though, to include contextualising premises before or after sentences to be recorded, on the grounds that it was likely to overwhelm the speaker with an excess of highly similar and repetitious utterances. Instead, a 'crib sheet' was produced that illustrated each context in pairs of sample premises, both presented using the typographical conventions described. In the result, however, the absence of a partner premise to each sentence in the script proved to be an obstacle to achieving the correct intonation. Consequently, in the later stages of recording, the speaker resorted to constructing such premises for himself 'on the fly'. It was here that his prior experience and knowledge of syllogisms and linguistics proved invaluable.

5.2.4 Voice recording

Clark *et al.* (2004) observes that the quality of the recording environment and equipment is not in fact a major concern in the building of a voice and that acceptable results can be obtained using relatively rudimentary facilities. Nonetheless, better results are got from better facilities and they were available for this project. The speaker was recorded in a sound-proofed, acoustically damped recording studio fitted with professional quality equipment. The recorded speech was saved directly to computer disk as a series of sampled data files in WAV format, each file containing batches of sentences of varying lengths.

5.2.5 Specification of linguistic resources

The computational phase of the voice building process requires a number of files specifying linguistic information to be provided. Included in these are an inventory of the phones used and their permitted replacements, a lexicon listing every vocabulary item and its pronunciation in the given phone set, and rules for determining phrase types and boundaries. For the syllogism voice, these were copies in their entirety of those used to build the COMIC voice. To these was added a file of lexical entries for the 24 terms, which did not feature in the domain of the earlier voice.

Also required is a set of files, one for each script sentence, containing the words of each recorded utterance marked up in APML for their intonation.[6] For present purposes, only the subset of the APML tagset that deals with intonation needed to be used. This is based on Steedman's (2004) theory of information structure and intonation, but this is not a bar to exploiting it. The relationship between the APML and multisyn modules in FESTIVAL is essentially 'theory-neutral', in that the former is used descriptively, not prescriptively. It will be recalled that the unit selection module does not generate the various elements of the linguistic signal themselves, but merely reproduces the prerecorded bundles of them that reside in the database, according to the selection criteria it is given. The APML module enables those criteria to include intonational markup. Provided the script sentences are both consistently marked up and consistently intoned, then, the specification of a given tag will be reliably associated with whatever intonation the speaker actually used for items so tagged, rather than whatever intonation any particular theory dictates ought to be used. Briefly put, in Humpty Dumpty style, the tags mean what the voice builder wants them to mean.

[6] Such files are also used to drive synthesis using the finished voice.

As with the FLIGHTS voice, the machinery required by Steedman's handling of mutual belief and contentiousness is not needed and only the H* and L+H* accents are adopted. The former is used to represent ordinary pitch accents, including focal accents, and the latter to represent exaggerated emphasis. Similarly, the only tone accents used are LH% and LL%, for continuation rise and standard falling tone, i.e. first and second premises, respectively. Following the decision made in respect of presentation of the script, intermediate tone boundaries were ignored.

5.2.6 Phone labelling

The phone-level labelling of the recorded utterances is a 'forced alignment' procedure consisting of a series of automated steps using the resources described above and a Hidden Markov Models toolkit. Its aim is to specify the boundaries between successive phones as time points in the sampled speech waveforms. In summary, this is achieved in two stages. First, the synthesiser, not in unit selection mode, is used to generate the correct sequence of phone labels for each utterance. Then, the toolkit is used to iteratively approximate the corresponding phone boundaries. The end result is a set of label files describing the phonetic segmentation of the speech database.

Join costs are calculated with reference to phone boundaries and therefore the accuracy of this phase is vital to the quality of the voice. On the syllogism set of utterances, the toolkit performed extremely poorly. In nearly every case, there were single phone labels spanning sequences of three or four recorded phones, followed by sequences of labels covering minimal spans. In very few cases were any phone boundaries accurately placed. The practical outcome of this was that the majority of phones in the database were placed effectively out of use and the only way of generating output that conformed to an input specification at all was to disable the join cost function completely, producing very low quality synthesis. Unfortunately, the models used in the HMM toolkit are fixed, so the only means of correcting these errors was to examine every one of the 696 pairs of label and waveform files individually and manually adjust the labels. The end result of this laboriously time-consuming task was a vast improvement in voice quality.

5.2.7 Utterance building

The final phase of database construction involves augmenting the labelled utterances produced in the previous phase with various levels of linguistic analysis. Once again,

the synthesiser is used to generate the phone sequence for each sentence in the script. This time, however, it builds on top of it the descriptions of the sentence's syllabic, grammatical, and other linguistic structures, known in FESTIVAL as 'relations'. Included in these are relations for its intonational make-up as given in its APML specification. The two phone sequences are then unified, producing the final database of recorded phones linked to a variety of interrelated structural annotations.

5.2.8 Specification of synthesis parameters

At run time, the only component of the system with a significant impact on voice quality that can be changed by the user is the target cost calculation. Join costs are precomputed according to a fixed formula, but target costs are computed on demand according to one of several formulae provided by the system, as selected by the user. Each of these is a weighted sum of normalised components representing linguistic features, such as lexical stress, position in syllable, word, and phrase, part of speech, preceding phonetic context, following phonetic context, and punctuation. Included in the set is one formula, intended for use in APML-based synthesis, which also contains components representing pitch accents and boundary tones. In tests, however, it proved unreliable, producing output that featured accents where none was specified, and vice versa, and incorrect boundary tones. Investigation of the problem revealed there to be two flaws in the calculation.

The primary fault lay in the two disjoint sets of relations for intonational marking supported by FESTIVAL. APML tags are recorded in the SemStructure, Emphasis, and Boundary relations, which register theme and rheme, pitch accents, and boundary tones, respectively. Each of these relates its tags to XML-derived tokens for text components in the Token relation and thereby, via the Word, SylStructure, and finally Segment relations, to individual phones. Alongside this, ToBI-style accents and tones are stored in the IntEvent relation. These are linked via the Intonation relation to syllables and thereby, via the SylStructure and Segment relations, again to phones.

The function of the latter set of relations is to register intonation events predicted by the system when synthesising in some mode other than unit selection mode. When synthesising directly from APML input in such a case, the prediction method is actually a forced selection or 'pass-through' of the accents and tones specified in the markup, resulting in the two sets of relations agreeing. In the unit selection voice building process, however, a different prediction method is used and the two are left out of

step. This ought not to matter, as the predicted set can simply be ignored in the target cost calculation.[7] Unfortunately, though, the implemented APML formula assumes the pass-through method has been used and tests only the predicted set of intonation events, leading to the selection of potentially inappropriate units.

A new, 'strict' APML target cost function was implemented to correct this fault. However, while it delivered output that conformed to the specification of pitch accents in the input, that output proved still to be defective in respect of boundary tones. Further investigation revealed that this was due to the accent and tone tests both being applied only to vowels. This is satisfactory in the case of pitch accents, as they are associated with syllables and therefore effectively with at least one vowel in each instance. Boundary tones, in contrast, occur at the very ends of words, which, in English, frequently consist of clusters of consonants. In fact, the pervasive presence of terms ending with '-ns' in the syllogism domain as defined was sufficient to render the test largely useless. In the final version of the cost function, then, it was extended to voiced consonants.

5.3 Conclusion

The completed voice demonstrably produces high quality synthesis that conforms to its input specification. The output utterances can be shown to consist overwhelmingly of substantial single-source phone sequences drawn from appropriate input utterances, which is the goal of unit selection. The voice is capable of delivering any of the desired combinations of terms, mood, figure, and intonation on demand. This amounts to $40 \times 24 \times 23 = 22080$ distinct premise utterances in total. Beyond this there is an indefinite number of utterances, of no doubt variable quality, that are ill-formed formally (e.g. "All Mexicans are not butchers."), grammatically (e.g. "no are hillwalkers no some."), semantically (e.g. "No golfers are golfers."), and intonationally (e.g. with no pitch accents at all).

The successful development of this voice, though, leaves one issue outstanding. Substantive corrective measures were required here in order to ensure that the synthesised output matched the input specification. Specifically, both the join and target costs

[7] There are other reasons why perhaps it ought not to matter, such as that the markup and the speaker's rendition of it are prescriptively accurate. However, the earlier discussion of APML's theory-neutral implementation explains why this can not be relied on. Furthermore, in the context of the present application, there are alternative intonations of identical wordings to accommodate, which text-based predictive methods can not deal with.

were shown to be severely compromised, requiring labelling and cost function correction. That being so, an explanation is required of why this was not so in the cases of the FLIGHTS and COMIC voices. Two are offered here.

First, both of the earlier projects focussed on the generation of *contextually appropriate* intonation in much more natural and contextualised utterances than those in the syllogism domain. This might have led the intonation events predicted by the synthesiser during the build process to approximate the specified contours quite well, in fact, leading to accidentally correct selection in the majority of cases. This applies equally to the 'inappropriate' ones tested in the FLIGHTS evaluation, which were appropriate to the contexts in which they were recorded and intended for and merely used out of context for evaluative purposes. Their contours were presumably significantly different from those of the utterances appropriate to the context of use and therefore capable of being distinguished from them by the synthesiser, but again not actually on the basis of the APML markup.

Second, both projects relied on user preference studies for evaluation of their output, rather than detailed analysis of the synthesised structures themselves. In such studies, participants do not have precisely detailed expectations of the intonation contours they hear and are merely required to state which ones they find more acceptable. In the FLIGHTS case, at least, the results as reported leave open the possibility that the synthesised utterances presented to participants did not conform exactly to the input specifications. Furthermore, only rather modest preferences were elicited there and not in all cases. Given that the original recordings received much higher ratings, this really ought not to be so if the synthesiser is reproducing the same contours faithfully.

Chapter 6

Syllogism solving using spoken materials

6.1 Introduction

To resume the primary course of this thesis, it will be recalled that the study reported earlier led to a rather surprising outcome. The intonation patterns actually used by participants were shown to be either neutral or concordant, i.e. based on the latent information structure of premises, and to be clearly related to the conclusions they drew, especially in respect of number and correctness, yet it remained unclear that they were instrumental in arriving at those conclusions. If Stenning and Cox's (2006) theory holds true here, then they should be, but the fact that individual participants seem only sometimes to respect the latent structure suggests otherwise, unless a further, as yet unknown, prior cause is at play.

The experiment described in this chapter addresses this question by imposing pre-specified intonation patterns on premises, using the synthetic voice introduced in the last chapter. The exercise is exactly the same as that given to participants in the standard conclusion generation task, save only that instead of reading the premises, they hear them spoken aloud. This might be thought to render an already quite difficult test inordinately so, but Gilhooly *et al.* (1993) have shown that the increase is but modest for aural as compared to visual presentation of premises.[1]

Table 6.1 summarises the degrees of correct and NVC responding to the subset of syllogisms in figures AB-CB and BA-BC with respect to the five major topicalisation

[1] In fact, their definition of correctness is unclear and leaves open the possibility that the increase in difficulty is less even than they suggest.

Topicalisation code		Conclusions		
		NVC ratio	Accuracy	Term ordering model?
Concordant:	C10	very high	very high	n/a
	T00	low	high	source founding
	T10	proportionate	proportionate	source founding
Neutral:	C01	proportionate	proportionate	recency
	T11	high	low	recency

Table 6.1: Characterisations of the five major topicalisation patterns in terms of three general properties of conclusions.

patterns, as discussed in Chapter 4. Extrapolating from this and the assumption that information structure does indeed play some part in the reasoning process, the following two hypotheses can be derived.

1. Syllogisms delivered with *CC* intonation patterns should generate more correct responses than those delivered with *CN* patterns (at least in figures AB-CB and BA-BC).

2. Syllogisms delivered with *CC* intonation patterns should generate fewer NVC responses than those delivered with *CN* patterns (at least in figures AB-CB and BA-BC).

The table also associates term ordering models with four of the five patterns, drawing on the speculation in Chapter 4 that Stenning and Cox's (2006) findings relate to a conception of grammar distinct from information structure properly so called. On this basis, a third hypothesis can be tentatively put forward.

3. Syllogisms delivered with *CC* intonation patterns should generate non-NVC responses whose terms are ordered according to the Source Founding Model, whereas those with *CN* patterns should generate such responses ordered according to recency (at least in figures AB-CB and BA-BC).

If the two main hypotheses are borne out, confirmation of this third would reconcile the different sets of findings.

6.2 Method

6.2.1 Materials

The primary materials used in the experiment were sound files synthesised on demand from APML specifications using the limited domain voice built for this purpose. The APML files were created in advance by a Prolog program similar to that used in the construction of the voice. In addition to these, a log file was automatically generated for each participant, listing the sequence of conditions used.

The program generated 64 such files per participant, constituting the full, standard set of syllogisms, half of which were marked up with *CC* contours and half with *CN*. Ideally, all 128 combinations of mood, figure, and intonation would have been presented to each participant, but this was considered excessively onerous. Each syllogism was instantiated with one of each of the three categories of terms (nationality, profession, and interest) at random and the ordering of the set of syllogisms was also randomised.

The two classes of intonation contours were distributed over the set of syllogisms according to one of three patterns. This was necessitated by the three orthogonal divisions into which the four quantifiers fall: positive vs. negative, universal vs. existential, and fully convertible vs. partially or non-convertible. Accounting for all of these requires three pairs of complementary contour assignments, as shown in Table 6.2, but these cannot be distributed equally across 64 instances. Consequently, the three patterns were formed by first dropping one of the complement pairs and then rotating the resultant square column-wise. Each of these was in turn rotated square-wise for each successive participant.

Quantifier	Polarity	Scope	Convertibility	Pair 1		Pair 2		Pair 3	
All	positive	universal	partial	*CC*	*CN*	*CC*	*CN*	*CC*	*CN*
Some	positive	existential	full	*CC*	*CN*	*CN*	*CC*	*CN*	*CC*
No	negative	universal	full	*CN*	*CC*	*CC*	*CN*	*CN*	*CC*
Some...not	negative	existential	none	*CN*	*CC*	*CN*	*CC*	*CC*	*CN*

Table 6.2: Pairs of complementary patterns of distribution of classes of intonation contours across the four quantifiers

6.2.2 Participants

The participants were 35 students drawn from across the University of Edinburgh. Each had responded to an advertisement placed in the university's student employment service and on the online forum run by the students' association. They were required to be native speakers of British English with no training in logic and no hearing impairments. Each was paid 5.05 for their participation.

The results of a further 6 participants were discarded for failure to satisfy the native language condition and yet another 7 for systematically ill-formed responses resulting from misinterpretation of the instructions.[2] Given the design of the experiment, the remaining 35 participants constituted 2 cohorts of 12 and 1 of 11, corresponding to the 3 intonation contour distribution patterns.

6.2.3 Procedure

The experiment was conducted remotely, via the World Wide Web. Since the standard task is essentially self-directed and self-paced, there seemed to be no reason why participants should not be given the facility to run the experiment for themselves at their convenience. It was considered that this would encourage take-up while reducing the administrative and logistical overheads of arranging to test significant numbers of participants. Participants were directed to a page on the World Wide Web which was dynamically generated by a PHP script. This displayed the instructions for the experiment and allowed the participant to verify their sound reproduction capability, then created a sequence of interactive forms corresponding to the 64 syllogisms, here called 'statement pairs'.

Each form constrained the participant to a single course of action. First, the participant was given a button for listening to the statement pair. Activating this button sent a request to the speech synthesiser to generate the utterance specified in the APML file for this participant and sequence number. This having been done, the participant was given a button for submitting a conclusion typed into a text entry box that was also provided. Activating this button validated and stored the conclusion, then generated the form for the next statement pair. At the end of the sequence, the script generated a data file combining the information listed in the participant's log file with the participant's responses, ready for analysis.

[2] In each of these cases, the entire response set consisted of instances of the following five statements: "All Xs are Ys", "Some Xs are Ys", "No Xs are Ys", "Some Xs are not Ys", "No valid conclusion".

Consideration was given to presenting the materials visually in tandem with the auditory presentation, to make the task more closely comparable to the standard version of it. In the end, however, technical and logistical considerations militated against this. With a parallel visual presentation arises the question of whether it should be static or windowed. A static presentation might enable participants to assist or bypass their auditory faculties using visuo-spatial strategies, while a windowed presentation raises technical issues of reading/hearing speed, fixation, and synchronisation. Additionally, there would have been a requirement for a control condition consisting of the visual presentation alone as well as the two auditory conditions. Given the structural factors already requiring to be taken into account, it was considered undesirable to increase the number of conditions further.

6.3 Results

6.3.1 Correctness and NVC-responding

Table 6.3 summarises the numbers of logically correct and NVC responses given by participants. The overall totals are within the normal bounds for the conclusion generation task, confirming Gilhooly *et al.*'s (1993) finding that the change in modality of presentation of the materials does not make it very much more difficult. The analysis by class of intonation contour reveals no significant difference in either accuracy or NVC-responding.

Contrary to hypothesis 1, *CN* intonation produced more correct conclusions than *CC* intonation. However, as the means and standard deviations indicate, the difference is not at all significant ($t = -1.09/F = 1.19, df = 34, p > 0.05$). In line with hypothesis 2, *CN* contours produced more NVC responses than *CC*, but only slightly and again

(a) Correct responses

	Total	%	Mean	SD
CC	455	40.6	13.00	6.42
CN	474	42.3	13.54	6.76
Total	929	41.5	13.27	6.55

(b) NVC responses

	Total	%	Mean	SD
CC	332	29.6	9.49	6.23
CN	335	29.9	9.57	6.43
Total	667	29.8	9.53	6.28

Table 6.3: Summary of frequencies of correct and NVC responses to syllogisms given concordant (*CC*) and neutral (*CN*) intonation patterns ($n = 35$)

(a) AB-BC

	Total	%	Mean	SD
CC	112	40.0	3.20	1.71
CN	118	42.1	3.37	2.04
Total	230	41.1	3.29	1.87

(b) BA-CB

	Total	%	Mean	SD
CC	111	39.6	3.17	1.89
CN	105	37.5	3.00	1.99
Total	216	38.6	3.09	1.92

(c) AB-CB

	Total	%	Mean	SD
CC	114	40.7	3.26	1.75
CN	124	44.3	3.54	1.80
Total	238	42.5	3.40	1.77

(d) BA-BC

	Total	%	Mean	SD
CC	118	42.1	3.37	1.93
CN	127	45.4	3.63	2.04
Total	245	43.8	3.50	1.98

Table 6.4: Summary of frequencies of correct responses to syllogisms in each of the four figures given concordant (*CC*) and neutral (*CN*) intonation patterns ($n = 35$)

the difference is in no way significant ($t = -0.19/F = 0.04, df = 34, p > 0.05$).

In Table 6.4, the results for correctness are broken down by figure. They are broadly in line with the figural effect on difficulty, both in the total frequencies and in the ones for the separate contour classes. Figure AB-BC presents a slight anomaly, in being the second hardest, after BA-CB, instead of the easiest. This perhaps represents an increase in the difficulty of problems with counterfigural conclusions brought about by sequential presentation. However, the main effect of figure approaches, but does not reach, significance ($F = 2.41, df = 102, p > 0.05$).

The general trend of the analysis by intonation contour matches the overall trend of *CN* generating marginally more correct answers than *CC*. Unsurprisingly, then, there is no significant interaction between contour and figure ($F = 0.53, df = 102, p > 0.05$). The exception to the trend is in figure BA-CB, where *CC* generates marginally more than *CN*. These results remain contrary to hypothesis 1, which postulated that *CC* should result in increased accuracy in the symmetrical figures, if nowhere else.

Table 6.5 shows the analysis by figure of NVC responses to *CC* and *CN* intonation. In line with earlier results, figure BA-BC generates more NVCs than AB-CB. Perhaps surprisingly, owing to the level of difficulty reported above, the other two figures produce the lowest numbers of NVCs. There is, however, no significant main effect of figure on NVC responding ($F = 1.26, df = 102, p > 0.05$).

With respect to the two intonation classes, figures AB-BC and AB-CB are near identical. In figure BA-CB*CN* gives fewer NVC responses than *CC*, while in figure BA-BCit

(a) AB-BC

	Total	%	Mean	SD
CC	78	27.9	2.23	1.82
CN	79	28.2	2.26	2.05
Total	157	28.0	2.24	1.92

(b) BA-CB

	Total	%	Mean	SD
CC	84	30.0	2.40	1.82
CN	76	27.1	2.17	1.81
Total	160	28.6	2.29	1.80

(c) AB-CB

	Total	%	Mean	SD
CC	85	30.4	2.43	1.84
CN	84	30.0	2.40	1.82
Total	169	30.2	2.41	1.81

(d) BA-BC

	Total	%	Mean	SD
CC	85	30.4	2.43	1.56
CN	96	34.3	2.74	1.84
Total	181	32.3	2.59	1.70

Table 6.5: Summary of frequencies of NVC responses to syllogisms in each of the four figures given concordant (*CC*) and neutral (*CN*) intonation patterns ($n = 35$)

produces more. It is this last that is responsible for the overall marginal compliance with hypothesis 2. The pattern is a curious one, being effectively orthogonal to trends in difficulty and the diagonal/symmetrical distinction, but the interaction is not significant ($F = 1.03, df = 102, p > 0.05$).

6.3.2 Term ordering

Table 6.6 presents the frequencies of well-formed, non-NVC responses according to the two term ordering models, source founding and recency. (In interpreting these results, it should be borne in mind that the predictions of the two models are not disjoint sets.) Recency is manifested in two thirds of the relevant responses, which is to be expected, as it incorporates the well-established and robust figural effect on term ordering in

(a) Source-founded responses

	Total	%	Mean	SD
CC	238	35.6	6.80	2.61
CN	251	38.2	7.17	2.12
Total	489	36.9	6.99	2.37

(b) Recency-based responses

	Total	%	Mean	SD
CC	433	64.8	12.37	4.45
CN	435	66.2	12.43	5.00
Total	868	65.5	12.40	4.70

Table 6.6: Summary of frequencies of source- and recency-based, well-formed, non-NVC responses to syllogisms given concordant (*CC*) and neutral (*CN*) intonation patterns ($n = 35$)

(a) AB-BC

	Total	%	Mean	SD
CC	68	38.2	1.94	1.03
CN	70	39.5	2.00	0.87
Total	138	38.9	1.97	0.95

(b) BA-CB

	Total	%	Mean	SD
CC	49	29.3	1.40	0.88
CN	51	31.7	1.46	0.85
Total	100	30.5	1.43	0.86

(c) AB-CB

	Total	%	Mean	SD
CC	54	33.5	1.54	1.04
CN	63	38.7	1.80	0.90
Total	117	36.1	1.67	0.97

(d) BA-BC

	Total	%	Mean	SD
CC	67	41.4	1.91	1.12
CN	67	42.9	1.91	1.15
Total	134	42.1	1.91	1.13

Table 6.7: Summary of frequencies of source-founded, well-formed, non-NVC responses to syllogisms in each of the four figures given concordant (*CC*) and neutral (*CN*) intonation patterns ($n = 35$)

figures AB-BC and BA-CB. Source-founding, on the other hand, accounts for just over a third of the responses, which is considerably lower than in Stenning and Yule (1997).

As with the results for correct and NVC conclusions, there is no significant main effect of intonation. Both classes of contour make near identical contributions to the numbers of recency-based conclusions, contrary to hypothesis 3, and therefore evidently have no significant effect there ($t = -0.10/F = 0.01, df = 34, p > 0.05$). Even further counter to the hypothesis, *CN* contours generate more source-founded responses than *CC*, but once again the difference is not significant ($t = -1.03/F = 1.07, df = 34, p > 0.05$).

The numbers of source-founded responses are analysed by figure in Table 6.7. The numbers remain low across the board, but another curious pattern, different from that for NVC responses, emerges. Figures AB-BC and BA-BC produce the largest numbers of source-founded conclusions, followed by AB-CB and then BA-CB, and this time the effect is significant ($F = 7.28, df = 102, p < 0.01$).

The relative frequencies of *CC* and *CN* intonation patterns, by contrast, show little variation. Only in figure AB-CB is there an appreciable difference between them, with *CN* generating more source-founded conclusions than *CC*, contrary to hypothesis 3. The interaction between contour and figure is not significant ($F = 0.21, df = 102, p > 0.05$).

Finally, Table 6.8 presents the figural analysis for recency-based conclusions. The

(a) AB-BC

	Total	%	Mean	SD
CC	144	80.9	4.11	2.11
CN	141	79.7	4.03	1.99
Total	285	80.3	4.07	2.04

(b) BA-CB

	Total	%	Mean	SD
CC	127	76.0	3.63	1.72
CN	124	77.0	3.54	1.74
Total	251	76.5	3.59	1.72

(c) AB-CB

	Total	%	Mean	SD
CC	65	40.4	1.86	1.26
CN	79	48.5	2.26	1.74
Total	144	44.4	2.06	1.52

(d) BA-BC

	Total	%	Mean	SD
CC	97	59.9	2.77	1.46
CN	91	58.3	2.60	1.75
Total	188	59.1	2.69	1.60

Table 6.8: Summary of frequencies of recency-based, well-formed, non-NVC responses to syllogisms in each of the four figures given concordant (*CC*) and neutral (*CN*) intonation patterns ($n = 35$)

totals are clearly weighted towards the established effect in figures AB-BC and BA-CB. Figure BA-BC also shows a similar, albeit reduced, tendency, while figure AB-CB is the only one to evidence a modest leaning away from recency-based responses. Again, the differences are large enough for the main effect of figure to be significant ($F = 21.05, df = 102, p = 0.00$).

The comparison between intonation contour classes essentially mirrors that for source-founding. In figure AB-CB, it is again *CN* patterns that feature more prominently, this time in line with the hypothesis, but in the other three figures there is, if anything at all, a slight preponderance of *CC* patterns. The interaction with figure is, once again, not significant ($F = 0.73, df = 102, p > 0.05$).

6.3.3 Other factors

Summaries of the interactions between intonation and mood are omitted as they add nothing material to the results already presented. That is, while they illustrate main effects for the quantifiers of both premises, there is no significant interaction between either of them and intonation.

6.4 Discussion

As they stand, these results completely refute the hypotheses extrapolated from the earlier study in the introduction to this chapter. No significant differences at all are seen to exist between the two classes of intonation contours on the measures used. In fact, the two can be seen to mirror each other right down to the level of individual participants. This extends to such minor disparities as do exist between these results and those of previous experiments. In all, performance of the task is effectively dominated by established figural effects, all the more so if the increased difficulty of figure AB-BC relative to the others is taken as an enhanced counter-figural effect.

This raises the question whether any effect that intonation might have is not simply being overwhelmed by those of figure, which are, after all, known to be powerful. This possibility can not be definitively discounted, but it should be remembered that the effects of figure do not for the most part actually reach significant levels in this experiment, although they do come close. That being so, the fact that the results for the two contour classes are so similar within each figure, including AB-BC, is surely telling. Of course, both of these facts could be accounted for as being in turn part of a floor effect induced by the increased difficulty of the task. However, it should also be remembered that the scores here, although a little depressed compared to the average, are still within normal bounds, just as in Gilhooly *et al.* (1993).

Alternatively, it is perhaps the case that initial differences in responses to the two classes of intonation patterns are lost over time. That is, allowing participants to take as long as they like to solve a syllogism enables them to work and rework the problem and thereby discard the import of intonation. If this is so, then it is hard to see how to overcome it. Experiments have been conducted in which participants were required to respond within a short time (e.g. Johnson-Laird and Bara, 1984), but combined with aural presentation this might well render the task too difficult. However, the real question that underlies this issue concerns the degree to which different components of working memory are involved in the reasoning process. The standard model of working memory divides it into three: the ‘visuo-spatial scratch-pad’, which enables spatial storage and rearrangement of limited amounts of visually-oriented input; the ‘phonological loop’, which enables sequential storage and rehearsal of limited amounts of aurally-oriented input; and the ‘central executive’, which maintains overall control over processing and allocation of resources (Baddeley, 1990). What little evidence there is on the respective involvement of these three components in syllogistic reason-

ing is far from conclusive.

Gilhooly *et al.*'s (1993) comparison of visual and verbal modes of presentation appears to indicate that the phonological loop is used only for initial coding and storage, as verbal presentation engenders only a modest increase in the number of errors made and this is comprised of a class of errors peculiar to it, in which conclusions contain the middle term. When suppression techniques were used, though, to load each of the three memory components separately throughout the reasoning task, they had a small but significant impact in respect of the phonological loop (albeit only of the reasoning task on the suppression task, not the other way round), whereas there was none at all in respect of the visuo-spatial scratch-pad. The only major effect produced was in respect of the central executive. In contrast, Quayle and Ball (2000) used tests of spatial and articulatory recall to classify participants prior to the reasoning task. Their results showed a significant difference in performance between high and low spatial recall ability, but none at all between high and low articulatory recall ability.

Both of these sets of findings must be treated with caution, though. The reasoning tasks used by Gilhooly *et al.* in conjunction with the suppression tasks and by Quayle and Ball were such that each syllogism was visually available to the participant throughout the reasoning process. This has no bearing on mental models accounts of reasoning, with which both studies were primarily concerned, as these nonetheless require some form of internalised representation and manipulation of the problem, but in general it fails to preclude the possibility that participants exploit the external representation of a problem directly when one is available and thereby largely obviate the need for working memory storage. This would certainly account for the findings of the former study, in which only the central executive was seen to play a major rôle.

As for the latter, it should also be noted that the two tests of recall ability, although drawn from prior published work in each case, were both visually presented and sequentially oriented. At the very least, this seems to blur the distinction between the two kinds of ability. In fact, whereas the articulatory recall test consisted of reproducing strings of letters presented all at once, the spatial recall test consisted of reproducing random temporal sequences of colour changes in a non-symmetrical array of boxes. This, taken together with the point concerning external representation, would render it unsurprising that it was those who scored highly in the latter that performed better in the reasoning task. On balance, then, it is suggested that, if anything, the weight of the evidence is in Gilhooly *et al.*'s favour and that the phonological loop plays an ongoing part in the reasoning process, one which might be amplified in the absence of

any continuing visual presentation of the problem.

Moving on from considerations relating to the nature of the task, then, there is the complementary possibility that the particular manipulation employed here is not powerful enough to produce effects of sufficient magnitude to be noticeable. That is, that merely *neutral* intonation does not stand in clear enough opposition to concordant intonation to produce an adequately large effect. True, a third class of contextually 'discordant' intonation patterns can be conceived, such as impose patterns of topicalisation and focussing that are directly at odds with the corresponding contextually concordant ones, and this would be an interesting avenue of exploration in further experiments. However, in the context of the present investigation, such patterns are not motivated either theoretically or by the observations made in the initial study. As such, it would have been difficult to derive detailed predictions concerning them and, in any case, to relate the results of using them to credulous and skeptical reasoning.

Another potential source of lack of power might be derived from an alternative characterisation of the increased difficulty of problems in figure AB-BC, namely that, overall, the symmetrical figures proved easier than the diagonal ones. This might suggest that the source of the change lies in the theoretical projections of intonation patterns produced in Chapter 3, or perhaps in their implementation in the artificial voice, on the ground that only projections for the symmetrical figures were empirically verified in the initial observational study. Further analysis of the speech corpus produced there and comparison with the output of the speech synthesiser are therefore required to confirm or deny this possibility.

Nonetheless, that effect is statistically non-significant. The only one that is not is that of figure (and mood) on term ordering. With respect to the recency model, it has already been indicated that the diagonal figures are largely responsible for this. The results for the other two are considerably closer to 50%. As for the source-founding model, the results perhaps represent a failure to replicate those of Stenning and Yule (1997). However, the purpose of this experiment was not to support or contradict the source-founding hypothesis *per se*, but, in part, to discover whether or not different classes of intonation pattern produce different degrees of compliance with it. As with accuracy and NVC responding, the pattern of results for *CC* and *CN* intonation match each other closely and no such difference can be found.

Taking all of the above into consideration, the issue of the power of the manipulation against the power of figure would seem to be the crucial obstacle in the way of a positive outcome here. Now, given that the effects of figure are known largely on the

basis of experiments carried out in the conclusion generation paradigm since its introduction by Johnson-Laird and Steedman (1978), a possible means of overcoming that obstacle is to employ one of the alternative paradigms of either conclusion selection or conclusion evaluation. Returning to the process/product distinction raised at the end of Chapter 4, it is argued that conclusion evaluation is the better choice, as the experimental materials would then resemble reports of completed syllogisms. The question then becomes whether or not manipulation of the intonation of such reports significantly affects judgements of their logical validity. The main drawback of this choice is, of course, that it would not address the issue of causation in respect of syllogism solving. However, a positive result from it would not only demonstrate that the intonation patterns that have been identified and observed are not merely redundant embellishments of finished products, it would also validate the projected intonation contours and the synthetic voice based on them and thereby bolster the outcome of the experiment in this chapter.

Chapter 7

Syllogism evaluation using spoken materials

7.1 Introduction

Even in the rarefied environment of the laboratory there exists the notion of a hearer for whom a speaker's utterances are intended, be it the technician, the researcher, or some other, more abstract personality. The results of the study presented in Chapter 4 leave open the possibility that the purpose of the intonation patterns observed there is to guide such a hearer's processing of completed syllogisms. By adopting the change in paradigm put forward at the end of the last chapter, the experimental participant is cast in the rôle of hearer and therefore this possibility can be investigated.

The experiment reported in this chapter, then, employs a variant of the conclusion evaluation task that is akin to the version of the conclusion generation task described in the previous chapter. As in that case, the only material departure from the standard task is that participants hear the syllogisms, now complete with conclusions, spoken aloud rather than read them from written sources.

The data from the initial study does not include evaluations of conclusions based on the spoken reports of them and so provides no empirical foundation on which to base a specific hypothesis to test in the present experiment. However, returning to the theoretical connection between different classes of intonation contours and credulous and skeptical reasoning does enable a hypothesis to be derived. Credulous reasoning assumes that there is a coherent and consistent model being conveyed, whereas skeptical reasoning does not and considers alternatives. Therefore, if *CC* intonation induces credulousness while *CN* induces skepticism, the former should lead to higher numbers

of conclusions being judged valid than the latter.

7.2 Method

7.2.1 Materials

As in the previous experiment, the primary materials used were sound files synthesised on demand from APML specifications using the purpose-built, limited domain synthetic voice. Again, a Prolog program was created to generate the APML and log files for each participant in advance. This time, though, the 64 APML files per participant specified the full, standard set of syllogisms together with conclusions. Selecting appropriate conclusions to use necessitated the resolution of two issues. The first was whether or not to present NVC conclusions as well as non-NVC conclusions and the second was how to control for the difference in difficulty of assessment of alternative conclusions to individual syllogisms.

If NVC conclusions are included, there is an imbalance in respect of available candidate conclusions between syllogisms for which NVC is the correct answer and those for which it is not. In the former case, any non-NVC conclusion is incorrect, whereas in the latter there is at least one correct non-NVC conclusion as well as multiple incorrect non-NVC conclusions and an incorrect NVC one. That is, when this is factored into validity, there are two possible conditions for NVC problems but three for non-NVC ones. If, on the other hand, only non-NVC conclusions are offered, then NVC problems are never presented with correct conclusions, again unbalancing the conditions. Indeed, participants may induce from this that there are no NVC problems in the problem space, potentially skewing their judgements.

Since the arguments generate some form of imbalance either way, this last point was the deciding factor, as it was considered important to maintain a conception of the problem space as similar as possible to those presented in the earlier tasks. Therefore, NVC conclusions were included, but not across the board. Since validity was one of the variables of primary interest, it was decided to restrict the conditions to one correct and one incorrect conclusion for both NVC and non-NVC problems. Presenting some incorrect NVC conclusions was necessary in order to avoid the possibility of participants inducing that it was always correct, but to do so in every case would have meant that no non-NVC problem was presented with both a correct and an incorrect non-NVC conclusion.

Of course, this meant that NVC conclusions also needed to be accounted for in terms of the second issue, that of matching the relative levels of difficulty of the correct and incorrect versions of each syllogism. There is very little in the way of guidance on this point unless a particular theoretical stance is taken, but it is obviously important to proceed on as empirical a footing as possible. Fortunately, Johnson-Laird and Steedman's (1978) study details, for every individual syllogism, not only the predictions of their model as it was formulated at that time, but also the frequencies, in terms of actual numbers of participants, with which different conclusions were drawn, including NVC. This, then, was used as a basis for selecting an incorrect conclusion with as similar a frequency to the correct one as the sparsity of data allowed.

The procedure adopted was as follows. For NVC problems, the non-NVC conclusion whose frequency was closest to that of NVC was chosen. For non-NVC problems, if at least one incorrect non-NVC conclusion was available, then again the one with the closest matching frequency was selected. If none was available, then NVC was selected. In a number of cases, the frequency match was not close, but was the best afforded by the data. In a few cases, a tie between competing candidates was decided arbitrarily. The resultant set contained 17 non-NVC problems with two non-NVC conclusions each and 10 with one NVC and one non-NVC conclusion. Of course, the remaining 37 were NVC problems with, again, one NVC and one non-NVC conclusion.

As in the previous experiment, half of the problems were marked up with *CC* intonation contours and the other half with *CN*, again distributed according to the three patterns used there. Additionally, though, equal numbers of correct and incorrect conclusions were distributed according to a complementary set of patterns in order to produce systematic variation of the two primary conditions, intonation and validity. Finally, as before, each syllogism was instantiated with one of each of the three categories of terms (nationality, profession, and interest) at random and the ordering of the set of syllogisms also randomised.

7.2.2 Participants

The participants were 24 students drawn from across the University of Edinburgh. Each had responded to an advertisement placed in the university's student employment service and on the online forum run by the students' association. They were required to be native speakers of British English with no training in logic and no hearing impairments. Each was paid 4.00 for their participation.

(a) Correct responses

	Total	%	Mean	SD
CC	433	56.4	18.04	3.67
CN	413	53.8	17.21	3.99
Total	846	55.1	17.63	3.81

(b) "Not valid" responses

	Total	%	Mean	SD
CC	337	43.9	14.04	2.14
CN	309	40.2	12.88	3.35
Total	646	42.1	13.46	2.84

Table 7.1: Summary of frequencies of correct and "Not valid" responses to syllogisms given concordant (*CC*) and neutral (*CN*) intonation patterns ($n = 24$)

7.2.3 Procedure

The experiment was again conducted via the World Wide Web. As previously, participants were directed to a page dynamically generated by a PHP script which presented the instructions for the experiment and gave the participant the opportunity to verify their sound reproduction capability. It then generated a sequence of interactive forms corresponding to the 64 syllogisms, this time referred to as 'arguments'.

The forms were very similar to those used in the conclusion generation experiment, each one limiting the participant's options to a single course of action. The participant was first given a button for listening to the argument. Clicking this button requested from the speech synthesiser the utterance corresponding to the description in the APML file for this participant and sequence number. The participant was then presented with a choice of two buttons, one labelled "Valid" and the other "Not valid", with which to register their evaluation of the argument they had just heard. Clicking either of these stored the appropriate response and then produced the form for the next argument. Having reached the end of the sequence, the script combined the information from the participant's log file with their responses to produce a single data file, ready for analysis.

7.3 Results

Table 7.1 summarises participants' evaluations with respect to logical correctness and the specific response "Not valid". These two are, of course, interdependent: "Not valid" is the correct response to invalid syllogisms and the incorrect response to valid ones. As explained in the introductory section, it is considered here as a potential 'Don't know' or 'Not sure' response, similar to the apparent use of NVC with *CN* intonation observed in the initial study.

(a) AB-BC

	Total	%	Mean	SD
CC	101	52.6	4.21	1.28
CN	98	51.0	4.08	1.59
Total	199	51.8	4.15	1.43

(b) BA-CB

	Total	%	Mean	SD
CC	92	47.9	3.83	1.49
CN	108	56.3	4.50	1.32
Total	200	52.1	4.17	1.43

(c) AB-CB

	Total	%	Mean	SD
CC	113	58.9	4.71	1.40
CN	98	51.0	4.08	1.64
Total	211	54.9	4.40	1.54

(d) BA-BC

	Total	%	Mean	SD
CC	127	66.1	5.29	1.04
CN	109	56.8	4.54	1.35
Total	236	61.5	4.92	1.25

Table 7.2: Summary of frequencies of correct responses to syllogisms in each of the four figures given concordant (*CC*) and neutral (*CN*) intonation patterns ($n = 24$)

(a) Valid syllogisms

	Total	%	Mean	SD
CC	240	62.5	10.00	1.89
CN	244	63.5	10.17	2.43
Total	484	63.0	10.08	2.15

(b) Invalid syllogisms

	Total	%	Mean	SD
CC	193	50.3	8.04	2.33
CN	169	44.0	7.04	2.77
Total	362	47.1	7.54	2.58

Table 7.3: Summary of frequencies of correct responses to valid and invalid syllogisms given concordant (*CC*) and neutral (*CN*) intonation patterns ($n = 24$)

Overall accuracy is higher than in the conclusion evaluation task and well within normal bounds. *CC* intonation patterns generate more correct responses than *CN* ones, but the difference is not significant ($t = 1.11/F = 1.23, df = 23, p > 0.05$). Intriguingly, they also produce more "Not valid" responses than *CN*, but again the difference is not significant ($t = 1.54/F = 2.38, df = 23, p > 0.05$). The more detailed analyses are illuminating on this point.

Tables 7.2 and 7.3 analyse correct responses by figure and validity of syllogism, respectively. There is a definite gradation of accuracy across the figures, with the two diagonal figures producing the fewest correct answers, followed by AB-CB, and then BA-BC.. This effect is significant ($F = 4.06, df = 69, p = 0.01$). Moreover, with the exception of BA-CB, this trend is mirrored in an increasing superiority of *CC* over *CN*intonation contours. The interaction between intonation and figure is also significant ($F = 3.93, df = 69, p < 0.05$).

(a) AB-BC

	Total	%	Mean	SD
CC	87	45.3	3.63	1.56
CN	72	37.5	3.00	1.79
Total	159	41.4	3.31	1.69

(b) BA-CB

	Total	%	Mean	SD
CC	92	47.9	3.83	1.24
CN	86	44.8	3.58	1.47
Total	178	46.4	3.71	1.35

(c) AB-CB

	Total	%	Mean	SD
CC	71	37.0	2.96	1.30
CN	88	45.8	3.67	1.71
Total	159	41.4	3.31	1.55

(d) BA-BC

	Total	%	Mean	SD
CC	87	45.3	3.63	1.21
CN	63	32.8	2.63	1.24
Total	150	39.1	3.13	1.31

Table 7.4: Summary of frequencies of "Not valid" responses to syllogisms in each of the four figures given concordant (*CC*) and neutral (*CN*) intonation patterns ($n = 24$)

(a) Valid syllogisms

	Total	%	Mean	SD
CC	144	37.5	6.00	1.89
CN	140	36.5	5.83	2.43
Total	284	37.0	5.92	2.15

(b) Invalid syllogisms

	Total	%	Mean	SD
CC	193	50.3	8.04	2.33
CN	169	44.0	7.04	2.77
Total	362	47.1	7.54	2.58

Table 7.5: Summary of frequencies of "Not valid" responses to valid and invalid syllogisms given concordant (*CC*) and neutral (*CN*) intonation patterns ($n = 24$)

In fact, *CC* and *CN* pattern very differently across the four figures in respect of accuracy. *CN* produces identical numbers of correct responses in figures AB-BC and AB-CBand near identical, but higher, numbers in figures BA-CB and BA-BC. *CC*, on the other hand, shows a steady increase in correctness from BA-CB up to BA-BC, by way of AB-BC and AB-CB, respectively.

Validity has a very evident impact an accuracy, with valid syllogisms generating significantly more correct answers than invalid ones ($F = 34.64, df = 23, p = 0.00$). Within that context, invalid syllogisms are judged correctly more often when given *CC* contours than when given *CN*, but intonation makes no difference to the evaluation of valid syllogisms and the interaction between intonation and validity is not significant ($F = 2.38, df = 23, p > 0.05$). There is, however, a significant three-way interaction between intonation, validity, and figure ($F = 2.97, df = 69, p < 0.05$).

In tables 7.4 and 7.5, the figural and validity-based analyses are repeated with re-

spect to responses of the form "Not valid". This time, figure BA-CB generates the highest number of responses and BA-BC the lowest, with the other two figures tied in between, but the effect is not significant ($F = 1.34, df = 69, p > 0.05$). In spite of this reverse, the increasing superiority of *CC* over *CN* follows the same general course as in the results for correctness, though the exception here is figure AB-CB, and the interaction between contour and figure remains significant ($F = 2.97, df = 69, p < 0.05$).

The patterns of results for the two classes of contours separately are both very different from each other and from those observed in the earlier tables. *CC* produces its fewest "Not valid" responses in figure AB-CB and its most in BA-CB, with figures AB-BC and BA-BC contributing equal numbers in between. *CN* progresses from its fewest in BA-BC to its most in AB-CB via AB-BC and then BA-CB.

Validity again has a significant effect, with invalid syllogisms gaining more "Not valid" responses than valid syllogisms ($F = 5.61, df = 23, p < 0.05$). This is not surprising, given the overall level of correctness and the fact that these figures are identical to those for correctness in respect of invalid syllogisms and the exact complements of them for valid ones. Also unsurprisingly, therefore, this minor change is not sufficient to render the interaction between contour and validity significant ($F = 1.23, df = 23, p > 0.05$). It does, however, lead to a significant interaction between validity and figure ($F = 4.06, df = 69, p = 0.01$), as well as a three-way interaction between contour, validity, and figure ($F = 3.93, df = 69, p < 0.05$).

7.4 Discussion

The results presented above refute the specific hypothesis derived at the beginning of this chapter, in that *CC* intonation contours show, albeit not to a significant degree, a greater tendency towards responding "Not Valid" than do *CN* contours. The effect of validity is powerfully evident across the board and serves to swamp the results somewhat, which, combined with the unusual figural interactions, makes them difficult to interpret. However, although no main effect of intonation has been shown, the existence of significant interactions between it and both validity and figure indicate that it does influence the evaluation of syllogisms.

Most interestingly, *CC* intonation generates more correct responses than *CN* does, but this is almost entirely accounted for by responses of the form "Not Valid" - the numbers of correct responses of the form "Valid" are about the same. That is, concordant intonation patterns appear to assist in the correct identification of invalid syl-

logisms, but make no difference to the identification of valid ones. If this is to be accommodated within a framework of credulous and skeptical reasoning, then it certainly needs to be a more sophisticated one than that used in the introduction to this chapter. A possibility is that the assumption of coherence implicit in credulous reasoning is contingent upon consistency with the implicatures licensed by such reasoning and breaks down when an inconsistent statement is encountered. In other words, a credulously processed syllogism is 'scaffolded' with all the other statements that are logically consistent with it, leading to a clash when an invalid conclusion is presented. Valid conclusions, by contrast, give rise to no such clash and so credulous reasoning affords no advantage over skeptical reasoning in respect of them.

However, this pattern of results is similar to certain interactions with validity known to exist and it is appropriate, therefore, to consider them in those lights. Atmosphere, as was mentioned in Chapter 2 (section 2.2.2.1), has a more pronounced effect in valid syllogisms than in invalid ones. This is the complete opposite of the pattern seen here and so the relevance of this effect to the present case is doubtful. Moreover, as Johnson-Laird and Steedman (1978) pointed out, atmosphere is incapable of accounting for the conclusion *No valid conclusion* under any circumstances. Instances of both valid and invalid conclusions of this type were included in the present experiment, further reducing the likelihood of these results being due to an interaction with atmosphere..

The influence of belief on syllogistic reasoning was touched on in the Introduction (Chapter 1). Studies of belief bias have shown, besides main effects for both the believability and the validity of conclusions, an interaction between the two, such that the effect of believability is greater for invalid syllogisms than for valid ones (see, generally Evans *et al.*, 1993). This is the same pattern as found for intonation here and so provides a more promising fit. There is no one agreed account of the logic-belief interaction; instead, explanations fall into three categories

The first of these posits the selective falsification of putative conclusions, such that the search for falsifying models is contingent upon the unbelievability of the conclusion. Some versions suggest that believable conclusions receive effectively no logical processing at all (e.g. Evans, 1989); others, including the orthodox mental models account (Oakhill and Johnson-Laird, 1985; Oakhill *et al.*, 1989), propose that all conclusions are generated from or tested against an initial, minimal model of the premises, but that only unbelievable ones are then subjected to further scrutiny. The second category of explanations proposes that participants misinterpret the notion of logical necessity and use believability as a fall-back heuristic in the face of uncertainty (Evans

et al., 1983; Quayle and Ball, 2000). That is, in the case of indeterminately invalid syllogisms, which admit of both confirmatory and disconfirmatory models, participants resolve the question on the basis of the believability or unbelievability of the conclusion. Lastly, there are accounts that suggest that a conclusion is only subjected to either confirmatory or disconfirmatory processing and that such processing is selected in the first place according to the believability or unbelievability, respectively, of the conclusion (Klauer *et al.*, 2000; Evans *et al.*, 2001).

Now, it would be easy to dismiss the first and third of these (selective falsification and selective processing) as inapplicable here, on the grounds that they fail to accommodate the results of the experiment reported in the previous chapter. The fact that some effect of intonation has been shown in the present chapter lends weight to the argument that those results constitute a genuinely negative outcome. that could not be the case, however, if intonation were allied to believability and believability determined some or all of the processing whereby a conclusion was reached. While this is indeed fatal to the standard mental models account, though, most of these theories assume that processing in the conclusion generation and conclusion evaluation paradigms proceed in opposite directions, the latter starting from the conclusion and working backwards, and therefore make no claims about the former.

However, none of these varieties of explanation accommodates the results of the initial observational study reported in Chapter 4. The pattern of usage observed there suggests that concordant intonation is employed when a participant arrives at a determination of the problem, whereas neutral is used when they are unsure of their answer or fail to reach one. The effect of believability, in any account, is to increase the likelihood of a conclusion being accepted, i.e. producing the response “Valid”, while that of unbelievability is to increase the likelihood of its rejection, i.e. of the response “Not Valid”. Therefore, according to the results in this chapter, *CC* intonation has the same effect as unbelievability and *CN* intonation has that of believability. That is, if there is a link between intonation and believability, then it is the former that induces doubt in the answer given, not the latter. On the face of it, then, this paints the curiously paradoxical picture of participants engaging in the self-defeating behaviour of casting doubt on the answers they are most sure of.

An additional reason for denying the applicability of belief bias to these results is that, like atmosphere, it does not extend to conclusions of the form *No valid conclusion*. As already mentioned, such conclusions were, however, included in the materials used in the present experiment. Moreover, the initial study in Chapter 4 demonstrated a

connection between the different classes of intonation contour and differential use of the response NVC. Indeed, variation in intonation there appeared to have as much to do with the validity of the response, whether NVC or otherwise, as with apparent certainty. This suggests an alternative interpretation of the results.

As has already been noted, studies of the logic-belief interaction have also consistently shown a significant effect of validity on its own, such that valid syllogisms are evaluated correctly more than invalid ones. This is strongly borne out by the present results: nearly two thirds of valid syllogisms in both intonation conditions were correctly evaluated, as against just over a third of neutrally intoned and half of concordantly intoned invalid ones. Perhaps, then, *CC* intonation does in some way assist in getting to the substance of the problem, but this is obscured by a ceiling effect in the case of the much easier valid syllogisms.

Finally, there are some more pragmatic possibilities that deserve to be noted. One is that these results are due to faulty implementation of the synthetic voice, such that its *CC* intonation is over-exaggerated and unnatural and thereby engenders doubt as to the genuineness of what is being said, while its *CN* intonation is more straightforward-sounding and therefore credible. Alternatively, it could be that the experimental setting or task itself is too artificial for contextually concordant intonation to be perceived as natural, whereas contextually neutral intonation is more appropriately ‘clinical’. However, neither of these on its own explains the restriction of the effect to invalid syllogisms and both resort again to believability, so the above discussion on that issue applies to them also.

Chapter 8

Conclusion

8.1 Summary and discussion of results

This thesis presents an innovative approach to the study of syllogistic reasoning and types of argumentative discourse. The experiments reported here are the first to attempt to capture, analyse, and control for intonational aspects of information structure in this context. They have produced materials of practical value to researchers wishing to mine this vein of investigation further and their results lead to a significant advance in understanding of discourse types and the rôle of information structure in them.

There are, in fact, three items of potential practical benefit to future researchers. The first is the corpus of recorded spoken responses to the conclusion generation task. Admittedly, it is only partially annotated at present, but even that portion demonstrably contains examples of both credulous and skeptical interpretations, while the remainder has at least been collected and awaits only labelling. Secondly, there is the systematic analysis of syllogistic premises in respect of latent information structure. Having been shown to be significant in both the production and comprehension of syllogisms, this is an aspect of the interaction of figure and mood to be considered and accounted for in future studies employing syllogisms. Third is the domain-specific, synthetic voice, which has been shown to be a valuable tool for conducting controlled experiments of the kind just adverted to.

Turning to the intellectual capital gained from the results reported here, it has been shown that information structure, as manifested in intonation, is a significant element in the both the presentation and comprehension of syllogisms. The initial study confirms that participants distinguish intonationally between sound and unsound arguments they have completed themselves. The two follow-up experiments show that this

distinction does not have the power to sway the completion of arguments one way or the other, but that it can be effective in mediating the assessment of unsound arguments. Characterised thus, these findings bear considerably on Stenning and Cox's (2006) account of interpreting and solving syllogisms, particularly its use of Stenning and van Lambalgen's (2005) distinction between expositive and deductive discourse types, as well as more broadly held conceptions of syllogisms.

Stenning and van Lambalgen highlight the importance of recognising the experimental situation itself as affording the potential for multiple, alternative interpretations of the type of discourse involved and, indeed, intended. Stenning and Cox apply this to syllogistic reasoning experiments in the specific context of the conclusion generation task, arguing that the problems presented in it can be construed in terms of either exposition or deduction, leading to different reactions to information packaging, and using individual differences in performance of the task as evidence of this occurring. In the present work, though, differences in information structure were seen to express differences in conclusion generation without producing them, but instead to influence conclusion evaluation.

One possible construal of this state of affairs is that Stenning and Cox's conception of information packaging is impoverished and their results are best interpreted purely at the syntactic level of problem grammar inducing solution grammar. If it is intonation, though, that gives the true guide to discourse types, then the distinction no longer appears to be that between exposition and deduction. Those two are both supposedly available as alternative interpretations at the problem-setting stage, giving rise to alternative forms of processing and conclusions. The distinction here, on the other hand, is between the syllogism presented as a problem and the syllogism presented as a solution, be it by the participant having completed it or by the experimenter for evaluation. Only in the latter case do intonation patterns appear to be significant.

Just what significance they have, though, is not clear. On the basis of the observational study, it is tempting to suggest that *CC* contours indicate what participants believe to be coherent proofs, whereas *CN* contours suggest dubious or inconclusive reasoning. Since there is an apparent association also with correctness, the purpose of such differential structuring would presumably be to assist in the interpretation of completed syllogisms as either sound or questionable, respectively, by whomever the participant conceives as being the recipient of their responses. However, if the results of the conclusion evaluation experiment are taken as symptomatic of believability, then these same contours have, if anything, the opposite effect from that intended. An alter-

native construal of these results is that *CC* intonation, by inducing credulous processing, allows the content of syllogisms to be 'scaffolded' with implicatures and thereby facilitates the detection of inconsistent, i.e. invalid, conclusions. As this is an explanation within the framework of Stenning and Cox's analysis of discourse types and processing, though, it is necessary to consider why these studies found no intonational correlate of Stenning and Cox's (2006) distinction.

The crucial difference between problem and solution in the conclusion generation task is, of course, the conclusion. This is not as facile an observation as it may at first seem, because, in the context of discourse types, it concerns the make-up of the whole. That is, a conclusion is not merely something added, or subtracted, but an integral part of a completed argument. Viewed from this perspective, the incomplete arguments presented as problems in the conclusion generation task are in a sense 'broken'. At the level of information structure, this might have the effect of neutralising the contextual component, through which relationships between statements are conveyed, and thereby rendering concordant intonation inapplicable. The lack of effect this class of intonation pattern has when applied to syllogisms requiring completion can therefore be put down to its being entirely inappropriate to that type of discourse.

In argument against this might be cited what Steedman (2000) calls 'isolated themes', or 'all-theme' utterances. As these appellations imply, these are considered not to convey any genuinely new information and therefore to be incomplete. Rather, the hearer is invited to work out the rheme for themselves precisely by way of their very marked intonation. This consists of a highly contrastive low-high pitch accent and a low-high boundary tone that is like a continuation rise, as in 8.1 (from Ward and Hirschberg, 1985; Pierrehumbert and Hirschberg, 1990). If intonation encourages the hearer towards drawing an appropriate conclusion in these cases of incompleteness, then the idea of its inapplicability to broken syllogistic arguments is questionable.

(8.1) A: Harry is such a klutz.
B: He's a good BADMINTON player.
L*+H LH%

There are several ways in which this objection might be countered. First, in implying contradiction of some part of what has gone before, the intonation pattern of these utterances is neither concordant nor neutral, as those terms have been used here. Concordant intonation, as its name suggests, fits in with and builds on what has gone before, while neutral delivers its content as though nothing has preceded it. Indeed,

contradictory intonation does not even fall within the class of 'discordant' contours mooted in Chapter 6 (section 6.4), as those would be manifestations of outright incoherence. (If a cognate term is desired, then 'contrapuntal' might fit the bill.) The fact that such intonation has inferential consequences, then, does not require the same to be true of other kinds. If, on the other hand, the argument is simply that *any* intonation should have inferential consequences, it is enough to note that neutral intonation is not the same as no intonation and this is therefore not the point of difference between it and concordant intonation..

Secondly, the analysis of such utterances as all-theme is by no means uncontroversial. It is necessitated in the context of Steedman's framework by the strict association asserted there between specific tunes and tones and specific informational constituents. Setting that aside, the fact that part of what is implied by such an utterance is that its content is already well known to the hearer is no bar to the pitch accent being analysed as focal. As was pointed out in Chapter 3 (section 3.1.1), the concept of 'newness' does not require the information to be completely new, either to the hearer or even to the current discourse. In 8.1, the fact of which B reminds A is, at that point, new to the issue A is considering, due to oversight, deliberate omission, or some other cause. These utterances, therefore, are not necessarily broken or incomplete at all.

Lastly, even if all-theme utterances are incomplete, they are not so in the same way that syllogisms as problems are. As Steedman (2000) observes, the former work as indirect speech acts, the reasoning triggered being in the nature of conversational implicature in order to complete the individual utterance around which they are drawn. They might, in other words, be considered as exhibiting a form of ellipsis. The fleshing out of the premises of a syllogism by way of implicature, on the other hand, is yet distinct from the coherent combination of them with an explicit conclusion. That is, none of the statements in a syllogism is incomplete in itself, intonationally or otherwise; it is the discourse as a connected whole that is 'broken' if the conclusion is missing.

Furthermore, this explanation accords quite well with the empirical data. First of all, although they were not used in the analysis, the responses of the four trial participants in the initial study are instructive. Delivered prior to the completion of the syllogism in each case, they are uniformly models of neutral intonation. Secondly, the main body of data from that study suggests that *CN* contours are associated with uncertain or inconclusive responses, i.e. ones that remain problematic.

If this is correct, though, it brings into question Stenning and Cox's use of different interpretations of the discourse type at the problem-setting stage to explain the differ-

ences in behaviour they found. Construing the problem as a kind of exposition supposedly induces credulous processing, but in that case *CC* intonation patterns would likely have appeared among the four excluded participants also. It is necessary to maintain the distinction between discourse types and processing modes, as well as bear in mind the potential for individual differences, in order to accommodate all of the phenomena that have been identified.

Even assuming that the problem itself and its associated rubric are strongly suggestive of deductive discourse, individuals nonetheless bring to it their personal biases towards credulity or skepticism. Notwithstanding these biases, though, either a credulous or a skeptical processor may generate either a sound or an unsound argument and, in the former case, thereby transform a problem into a solution. In this way, a discourse type essentially lacking a contextual dimension can be processed differentially and the outcome thereof manifested largely in the grammatical confines of the conclusion, yet the resultant whole be expressed simply according to whether or not a context is now seen to be present.

This resolution of the two previously seemingly incompatible views of information packaging and discourse types also embraces the 'scaffolding' explanation offered for the results of the final experiment. In melding the conceptualisations of the two, the discourse types originally proposed as being operative at the problem-setting stage, namely Stenning and van Lambalgen's (2005) exposition and deduction, have been mapped onto the ones claimed to be operative at the solution-giving stage. This is in spite of exposition being supposedly misappropriated to this context in the first place through the mistaken application of conversational assumptions. Moreover, since it contains the contextual connections that generate the concordant intonation contours that have been shown to be associated with them, it is this discourse type that is taken to be assigned to sound arguments, while deduction and its neutral contours are given over to unsound arguments.

The reason for this apparently topsy-turvy state of affairs is that intonation is, by itself, a relatively blunt instrument whose potential for signification is very limited. What the two concordant cases have in common is the relation of logical dependence between statements. The precise nature and number of implicatures this warrants will depend on the discourse context and its associated contextualised logic. In fact, then, the only genuinely incongruous aspect of this here is the name, 'exposition', which might therefore be replace with, say, 'demonstration'. Similarly, the neutral cases share the relation of logical independence and might be better termed 'examination'

at the problem-setting stage and, it is suggested more hesitantly, 'submission' at the responding stage.

In this light, it should be noted that Stenning and Cox's fundamental point that, contrary to Newstead (1989, 1995), interpretation does play a part in syllogistic reasoning, is supported and, indeed, extended. The combined results reveal interpretational effects in the processing of both examinations and submissions, even if not demonstrations. In fact, an even larger point can be argued. Aristotle's logic of the syllogism revolved around demonstration, not examination, and its relationship to its different possible modes of expression, referred to by the *rhetorical* term 'figure' (in Greek, actually σχημα, 'schema'). Completing 'broken' syllogisms, then, is perhaps a strange thing to ask people to do and one to which they may well bring more than logic. Moreover, though, given the pervasive influence of figure that has been shown in countless experiments and its interactions with validity and, as is now know, information structure, it seems reasonable to say that the human capacity exposed by studies of syllogistic reasoning is not that for logic alone but that for combining all three of the ancient *trivium* of logic, grammar, and rhetoric.

The major drawback of framing the discussion in terms of discourse types, however, is that it over-emphasises pragmatics at the expense of semantics. Participants' reactions to the experimental setting, their interpretations of the task, their notional interactions with the speaker/hearer, and so on are all aspects of context more to do with the discourse situation than the discourse content. True, the idea of 'scaffolding' - indeed, of implicature drawing in general - concerns the augmentation of some model of the propositional content conveyed by statements, but attention has not yet been paid to the grounding of this in a conception of the world of discourse. This is important because it harks back to the core difference between the classes of intonation contour studied here and their relationship to the notion of logical independence.

The fundamental distinction between concordant and neutral intonation was stated to be the inclusion or exclusion, respectively, of the contextual element of information structure (see Chapter 3, section 3.2.1). In syllogisms this is minimal, limited as it is to three sets of entities which, standardly, are either completely abstract or at least semantically unrelated and are embedded in a highly underspecified scenario. It is possible that this is not enough for the contrasts expressed by concordant intonation to be genuinely meaningful and so for the ability of such intonation to imply coherence or logical dependence is largely lost. This is an alternative explanation for the failure to produce stronger effects using the synthetic voice. The fact that *CC* intonation was

found in the initial study could be due to participants constructing a sufficiently rich representation of the semantic space in the course of reaching their conclusions.

Of course, the use of semantically richer materials can give rise to the sorts of difficulties that were mentioned in the introduction to this thesis. The most notable of these, though, is belief bias, which has already been put forward as a potential explanation for the results in Chapter 7. Although significant doubts were raised about its applicability there, the possibility of a more general connection between it and information structure by way of rich contextualisation of the problem is not to be dismissed out of hand. Indeed, it would be making a virtue of necessity to use explicitly biased materials to investigate the effects of intonation and believability in combination.

Looked at this way, and bearing in mind the link with validity seen in both the initial study and the conclusion evaluation experiment, intonation and information structure are best seen as logical factors according to the three-way classification of factors influencing syllogistic reasoning presented in Chapter 2, section 2.2.2. The discourse types view is much more integrative, drawing in both structural factors, such as figure and grammar, and individual differences as to interpretation and reasoning. Indeed, the division of these factors into structural, logical, and individual now appears to correspond at least crudely to the linguistic distinctions between syntax, semantics, and pragmatics.

Pursuing this idea further, though, belief can be separated into semantic and pragmatic components. As indicated above, the semantic component concerns the degree of congruence between the discourse situation and a participant's own prior knowledge and understanding of such situations. The pragmatic component, on the other hand, concerns the degree of confidence in or commitment to the coherence of the discourse as a legitimate progression of statements. Arguably, it is the latter which is reflected in the differential use of *CC* and *CN* intonation demonstrated in the observational study. If that is so, then it is no bar to the relevance of the logic-belief interaction to the results of the conclusion evaluation experiment after all, rooted as that interaction is in the semantic component of belief.

However, against this must be set the fact that the materials used in that experiment were designed to be belief-neutral. It is not easy to see how intonation geared to the pragmatic expression of coherence could colour the content of those materials to such an extent as to trigger a biasing effect, at least not in a direction opposite to the one pragmatically intended. Perhaps, then, the influence is the other way round, with concordant intonation being made to seem like over-compensation in the face of the

lack of semantic substance and therefore insincere.

In the light of these considerations, it is unfortunate that the analytical approach taken in this thesis affords no basis for distinguishing between these two aspects of belief. With hindsight, it might have proved more fruitful to have adopted Steedman's (2000) framework, which incorporates a third dimension of information structure that reflects differences between speaker and hearer in individual commitment to and agreement upon the contents of particular constituents. These are expressed by various combinations of specific pitch accents and phrasal boundary tones. It is perhaps through these extra subtleties, which of course were not controlled for in the building of the synthetic voice, that unbelievability can be generated from indifferent content.

Turning briefly, and finally, to the other interaction with validity considered in Chapter 7, atmosphere can probably be discounted as a distinct contributing factor to the results of the present work. Even leaving aside the poor fit of that interaction with the one observed in the conclusion evaluation experiment and the inapplicability of atmosphere to NVC conclusions, it seems likely that it is essentially simply a validity effect in disguise. As Johnson-Laird and Steedman (1978) observe, the majority of valid conclusions happen to be consistent with atmosphere. Unless experimenters are careful to choose an equal representation of atmospheric but invalid conclusions, which historically they have not, then atmosphere will accidentally pattern with validity more in valid syllogisms than invalid ones. Given this and the point concerning NVC conclusions already made, it has been argued that atmosphere is best considered merely descriptive, and only partially so at that, rather than explanatory (Eysenck and Keane, 1995). This leaves validity alone as the relevant factor, whose connection with intonation has been shown independently here.

8.2 Further work

8.2.1 Retrospective

There are some pieces of consolidatory work that could usefully be done. To begin with, the remainder of the corpus of recorded responses gathered in the initial study needs to be analysed and annotated. Indeed, the part that has already been analysed could be rechecked for consistency during this process. Besides improving the corpus as a resource, this would exhaustively confirm or refute the details of the theoretical informational analysis.

Secondly, it is possible that the synthetic voice could be improved. This would, of course, be required if the full corpus annotation led to revisions of the analysis. Even in the absence of that, though, the knowledge and experience gained from building the first version could feed forward to create a higher quality one. For example, re-designing the recording script to embed *CC* contexts in pairs of premises or completed syllogisms ought to produce more natural and convincing output from the speaker.

Thirdly, previous experiments could be replicated with the addition of the artificial voice in order to determine the reproducibility of their results with the added factor of intonation controlled for. Earlier experiments using the conclusion evaluation task would certainly be candidates for this, the potential being there for further light to be shed on some of the more well-established effects, such as conversion. For that matter, applying it to the immediate inference task might uncover hitherto unnoticed relationships. Conclusion generation experiments ought not to be overlooked, however. The theory and data from the present work indicate that their results should remain unaffected by the use of the voice, but this ought to be verified.

8.2.2 Prospective

Having reason to respect the importance of the conclusion as an integrated part of an argument, as opposed to an answer distinct from a question, it could be instructive to present both of these experiments in a more explicitly discourse-oriented light. The conclusion generation task, for example, could be recast as an argument completion task, the question being along the lines of, “What is the voice going to say next?” The idea behind this is that one aspect of a well-formed argument is that the intonationally signalled information structure anticipates the conclusion. In other words, in a more normal, communicative context, the conclusion is what the producer of the argument wants it to be, rather than necessarily the logically correct one, and it is this that the receiver looks for.

Similar considerations apply to the evaluation task and, indeed, provide an alternative explanation that should be investigated for the modest results of the experiment reported in Chapter 7. There is some anecdotal support for this idea from one of the participants in that experiment. He stated afterwards that he found the task hard because the voice was “too fast”, in spite of the fact that the speech output is quite measured. His subsequent elaboration of the point, however, revealed that he wanted a pause after hearing the two premises, during which he could form his own conclusion,

before then hearing the conclusion to evaluate. Hearing it straight after the premises interfered with his processing.

The instructions for the task were virtually identical to those for the generation task before it, i.e. heavily oriented towards problem-solving, with assumption of the premises and evaluation of the conclusion presented as distinct steps to take. Arguably, then, they were simply not appropriate to the discourse type actually presented. Had the instructions been, say, to indicate acceptance or rejection of the speaker's argument, it is possible that a different pattern of results would have been obtained.

These ideas in turn generate a fairly obvious line of consequent investigation, which is to probe the informational aspects of conclusions. The studies reported here concentrated mainly on the intonational expression of information structure in second premises, because that was where the clearest and widest variation was found. However, since the importance of including the conclusion as part of the argument has been shown, it should now receive the same attention.

The observational study has already shown that conclusions, like first premises, are overwhelmingly rheme-focussed, so the next step is to investigate the nature of topicalisation in them. Theory is not so clear a guide here, as both terms are given and, more often than not, so is the quantifier. It would be interesting, for that reason, to analyse responses in the speech corpus to the minority of syllogisms whose valid conclusion introduces a new quantifier, as well as seeking out instances of participants doing this erroneously.

Obviously, though, the aim would not simply be to characterise conclusions in themselves, but to relate their information structuring to that of the preceding premises. If conclusions are anticipated by their premises in some way, then one might expect to see topicalisation in the latter systematically related to patterns in the former. A useful additional light on this might be gained from what might be called 'B-conclusions'. these are the significant minority of ill-formed conclusions which feature the middle term instead of one of the end terms. Assuming investigations such as these produced useful results, more sophisticated conclusion evaluation experiments could then be run using the artificial voice.

To finish on an alternative note, a parallel line of investigation might examine the effects of orthography instead of intonation. It will be recalled that bold and italic fonts were used to represent accenting in the recording script for building the synthetic voice. The differential effects of varying these and other devices, such as capitalisation and underlining, in experimental syllogistic reasoning tasks would be interesting to

compare with those from the intonationally-oriented track.

Appendix A

Sample materials from initial prosodic analysis

A.1 Screen capture of annotation using `xwaves`

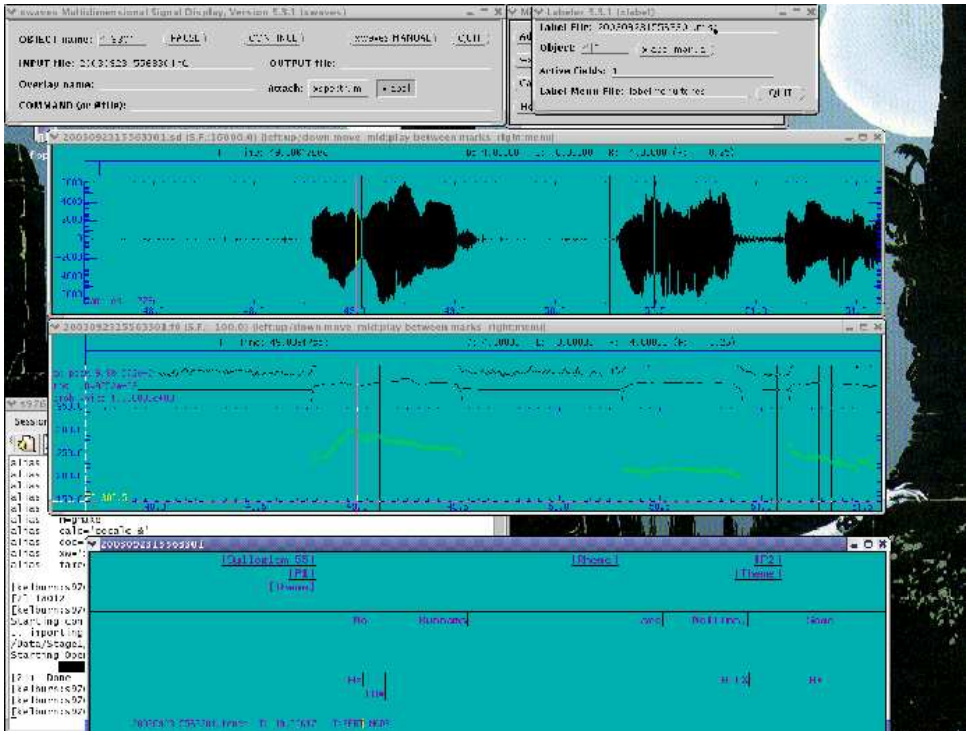

Figure A.1: Typical ensemble of windows used for annotation. Upper windows (grey) are for controls, lower windows (green) are for data. Data windows show aligned sections of the sampled speech waveform (top), the computed f0 contour (middle), and the label files (bottom) for the miscellaneous, word, and tone layers, respectively.

A.2 Sample portions of annotation files

These are for illustrative purposes only. Triple dots (...) indicate ellipsis.

A.2.1 Miscellaneous layer

```
signal 2003092315563301
type 0
comment created using xlabel Mon Mar 28 14:59:51 2005
comment created using xlabel Tue Jan  6 10:17:28 2004
font -misc-*-bold-*-*-*-15-*-*-*-*-*-*-*
separator ;
nfields 1
#
   48.773863   -1 [Syllogism 55]
   48.775780   -1 [P1]
   48.775780   -1 [Theme]
   50.275890   -1 [Rheme]
   51.085302   -1 [P2]
   51.085302   -1 [Theme]
   51.873129   -1 [Rheme]
   52.802408   -1 [Conc]
            ...
  424.586116   -1 [Syllogism 35]
  424.588027   -1 [P1]
  424.588027   -1 [Theme]
  425.409457   -1 [Rheme]
  426.252316   -1 [P2]
  426.252316   -1 [Theme]
  427.823747   -1 [Rheme]
  430.002323   -1 [Conc]
  430.002323   -1 [Theme]
  430.923753   -1 [Rheme]
            ...
```

A.2.2 Word layer

```
signal 2003092315563301
type 0
comment created using xlabel Thu Feb  5 15:25:39 2004
comment created using xlabel Tue Jan  6 10:17:41 2004
font -misc-*-bold-*-*-*-15-*-*-*-*-*-*-*
separator ;
nfields 1
#
   49.033676   -1 No
   49.524433   -1 Runnams
   50.494401   -1 are
   50.915875   -1 Mellins.
   51.348896   -1 Some
   51.781918   -1 Runnams
   52.053278   -1 are
   52.555582   -1 Norrels.
   53.040566   -1 No
   53.363889   -1 valid
   54.086517   -1 conclusion.
             ...
  424.862083   -1 No
  425.404851   -1 Larrins
  425.525466   -1 are
  426.121842   -1 Nemmins.
  426.503790   -1 All
  426.966148   -1 Murrons
  428.145498   -1 are
  429.097018   -1 Nemmins.
  430.356775   -1 No
  430.852638   -1 Larrins
  431.100569   -1 are
  431.603133   -1 Murrons.
             ...
```

A.2.3 Tone layer

```
signal 2003092315563301
type 0
comment created using xlabel Tue Apr 26 16:37:08 2005
comment created using xlabel Tue Jan  6 10:17:28 2004
font -misc-*-bold-*-*-*-15-*-*-*-*-*-*-*
separator ;
nfields 1
#
   49.006174   -1 H*
   49.118222   -1 !H*
   50.924010   -1 H-L%
   51.293434   -1 H*
   52.144119   -1 !H*
   52.565914   -1 L-L%
   53.027148   -1 H*
   53.841219   -1 !H*
   54.099826   -1 L-L%
            ...
  425.109457   -1 H*
  425.609458   -1 !H*
  426.130416   -1 H-L%
  426.473745   -1 H*
  426.588031   -1 !H*
  428.602320   -1 !H*
  429.105034   -1 L-L%
  430.323401   -1 H*
  430.459259   -1 !H*
  431.195159   -1 !H*
  431.619815   -1 L-L%
            ...
```

A.2.4 Combined, formatted annotation file

```
[Syllogism 35]  [P1]    [Th]            No
                                H*      Larrins
                        [Rh]            are
                        [f->]   !H*     Nemmins.
                                H-L%
                [P2]    [Th]    H*      All
                                !H*     Murrons
                        [Rh]            are
                        [f->]   !H*     Nemmins.
                                L-L%
                [Conc]  [Th]    H*      No
                                !H*     Larrins
                        [Rh]            are
                        [f->]   !H*     Murrons.
                                L-L%
                     ...
[Syllogism 55]  [P1]    [Th]    H*      No
                        [f->]   !H*     Runnams
                        [Rh]            are
                                        Mellins.
                                H-L%
                [P2]    [Th]    H*      Some
                                        Runnams
                        [Rh]            are
                        [f->]   !H*     Norrels.
                                L-L%
                [Conc]          H*      No
                                        valid
                        [f->]   !H*     conclusion.
                                L-L%
                     ...
```

Appendix B

Prolog data used in voice synthesis experiments

B.1 Patterns for distribution of intonation and validity conditions across the 64 syllogisms

```
% Intonation contour/validity groupings

cvgroup(0, [cvset(a1,ac1), cvset(b1,bc1), cvset(a2,ac2), cvset(b2,bc2)]).
cvgroup(1, [cvset(a1,ab1), cvset(c1,bc1), cvset(a2,ab2), cvset(c2,bc2)]).
cvgroup(2, [cvset(b1,ab1), cvset(c1,ac1), cvset(b2,ab2), cvset(c2,ac2)]).

contourset(a1, [neutral, neutral, concordant, concordant]).
contourset(a2, [concordant, concordant, neutral, neutral]).
contourset(b1, [neutral, concordant, concordant, neutral]).
contourset(b2, [concordant, neutral, neutral, concordant]).
contourset(c1, [neutral, concordant, neutral, concordant]).
contourset(c2, [concordant, neutral, concordant, neutral]).

validityset(ab1, [valid, invalid, valid, invalid]).
validityset(ab2, [invalid, valid, invalid, valid]).
validityset(ac1, [valid, invalid, invalid, valid]).
validityset(ac2, [invalid, valid, valid, invalid]).
validityset(bc1, [valid, valid, invalid, invalid]).
validityset(bc2, [invalid, invalid, valid, valid]).
```

B.2 Specifications of intonation contours derived fro mlatent information structure of syllogisms

```
% Intonation contours
% contour(Premise, Figure, Quantifiers, Type, Accents).

contour(1, _, [a, _], _, [hstar, hstar, none, hstar, lh]).
contour(1, _, [i, _], _, [hstar, hstar, none, hstar, lh]).
contour(1, _, [e, _], _, [hstar, hstar, none, hstar, lh]).
contour(1, _, [o, _], _, [hstar, hstar, none, none, hstar, lh]).

contour(2, _, [_, a], neutral, [hstar, hstar, none, hstar, ll]).
contour(2, _, [_, i], neutral, [hstar, hstar, none, hstar, ll]).
contour(2, _, [_, e], neutral, [hstar, hstar, none, hstar, ll]).
contour(2, _, [_, o], neutral, [hstar, hstar, none, none, hstar, ll]).

contour(2, 1, [a, a], concordant, [none, hstar, none, hstar, ll]).
contour(2, 1, [i, a], concordant, [hstar, hstar, none, hstar, ll]).
contour(2, 1, [e, a], concordant, [hstar, hstar, none, hstar, ll]).
contour(2, 1, [o, a], concordant, [hstar, hstar, none, hstar, ll]).

contour(2, 2, [a, a], concordant, [none, lplushstar, none, hstar, ll]).
contour(2, 2, [i, a], concordant, [hstar, lplushstar, none, hstar, ll]).
contour(2, 2, [e, a], concordant, [hstar, lplushstar, none, hstar, ll]).
contour(2, 2, [o, a], concordant, [hstar, lplushstar, none, hstar, ll]).

contour(2, 3, [a, a], concordant, [none, hstar, none, none, ll]).
contour(2, 3, [i, a], concordant, [hstar, hstar, none, none, ll]).
contour(2, 3, [e, a], concordant, [hstar, hstar, none, none, ll]).
contour(2, 3, [o, a], concordant, [hstar, hstar, none, none, ll]).

contour(2, 4, [a, a], concordant, [none, none, none, hstar, ll]).
contour(2, 4, [i, a], concordant, [hstar, none, none, hstar, ll]).
contour(2, 4, [e, a], concordant, [hstar, none, none, hstar, ll]).
contour(2, 4, [o, a], concordant, [hstar, none, none, hstar, ll]).
```

```
contour(2, 1, [a, i], concordant, [hstar, hstar, none, hstar, ll]).
contour(2, 1, [i, i], concordant, [none, hstar, none, hstar, ll]).
contour(2, 1, [e, i], concordant, [hstar, hstar, none, hstar, ll]).
contour(2, 1, [o, i], concordant, [none, hstar, none, hstar, ll]).

contour(2, 2, [a, i], concordant, [hstar, lplushstar, none, hstar, ll]).
contour(2, 2, [i, i], concordant, [none, lplushstar, none, hstar, ll]).
contour(2, 2, [e, i], concordant, [hstar, lplushstar, none, hstar, ll]).
contour(2, 2, [o, i], concordant, [none, lplushstar, none, hstar, ll]).

contour(2, 3, [a, i], concordant, [hstar, hstar, none, none, ll]).
contour(2, 3, [i, i], concordant, [none, hstar, none, none, ll]).
contour(2, 3, [e, i], concordant, [hstar, hstar, none, none, ll]).
contour(2, 3, [o, i], concordant, [none, hstar, none, none, ll]).

contour(2, 4, [a, i], concordant, [hstar, none, none, hstar, ll]).
contour(2, 4, [i, i], concordant, [none, none, none, hstar, ll]).
contour(2, 4, [e, i], concordant, [hstar, none, none, hstar, ll]).
contour(2, 4, [o, i], concordant, [none, none, none, hstar, ll]).

contour(2, 1, [a, e], concordant, [hstar, hstar, none, hstar, ll]).
contour(2, 1, [i, e], concordant, [hstar, hstar, none, hstar, ll]).
contour(2, 1, [e, e], concordant, [none, hstar, none, hstar, ll]).
contour(2, 1, [o, e], concordant, [hstar, hstar, none, hstar, ll]).

contour(2, 2, [a, e], concordant, [hstar, lplushstar, none, hstar, ll]).
contour(2, 2, [i, e], concordant, [hstar, lplushstar, none, hstar, ll]).
contour(2, 2, [e, e], concordant, [none, lplushstar, none, hstar, ll]).
contour(2, 2, [o, e], concordant, [hstar, lplushstar, none, hstar, ll]).

contour(2, 3, [a, e], concordant, [hstar, hstar, none, none, ll]).
contour(2, 3, [i, e], concordant, [hstar, hstar, none, none, ll]).
contour(2, 3, [e, e], concordant, [none, hstar, none, none, ll]).
contour(2, 3, [o, e], concordant, [hstar, hstar, none, none, ll]).
```

```
contour(2, 4, [a, e], concordant, [hstar, none, none, hstar, ll]).
contour(2, 4, [i, e], concordant, [hstar, none, none, hstar, ll]).
contour(2, 4, [e, e], concordant, [none, none, none, hstar, ll]).
contour(2, 4, [o, e], concordant, [hstar, none, none, hstar, ll]).

contour(2, 1, [a, o], concordant, [hstar, hstar, none, hstar, hstar, ll]).
contour(2, 1, [i, o], concordant, [none, hstar, none, hstar, hstar, ll]).
contour(2, 1, [e, o], concordant, [hstar, hstar, none, hstar, hstar, ll]).
contour(2, 1, [o, o], concordant, [none, hstar, none, none, hstar, ll]).

contour(2, 2, [a, o], concordant, [hstar, lplushstar, none, hstar, hstar, ll]).
contour(2, 2, [i, o], concordant, [none, lplushstar, none, hstar, hstar, ll]).
contour(2, 2, [e, o], concordant, [hstar, lplushstar, none, hstar, hstar, ll]).
contour(2, 2, [o, o], concordant, [none, lplushstar, none, none, hstar, ll]).

contour(2, 3, [a, o], concordant, [hstar, hstar, none, hstar, none, ll]).
contour(2, 3, [i, o], concordant, [none, hstar, none, hstar, none, ll]).
contour(2, 3, [e, o], concordant, [hstar, hstar, none, hstar, none, ll]).
contour(2, 3, [o, o], concordant, [none, hstar, none, none, none, ll]).

contour(2, 4, [a, o], concordant, [hstar, none, none, hstar, hstar, ll]).
contour(2, 4, [i, o], concordant, [none, none, none, hstar, hstar, ll]).
contour(2, 4, [e, o], concordant, [hstar, none, none, hstar, hstar, ll]).
contour(2, 4, [o, o], concordant, [none, none, none, none, hstar, ll]).
```

B.3 Conclusions taken from Johnson-Laird and Steedman (1978) for use in Experiment 3

```
% Conclusions
conc(n, _, ['Therefore,', 'no', 'valid', 'conclusion']).
conc(a, terms(A, C), ['Therefore,', 'all', A, 'are', C]).
conc(i, terms(A, C), ['Therefore,', 'some', A, 'are', C]).
conc(e, terms(A, C), ['Therefore,', 'no', A, 'are', C]).
conc(o, terms(A, C), ['Therefore,', 'some', A, 'are', 'not', C]).

% conclusion(Figure, Quantifiers, Type, Terms, Conclusion).
conclusion(1, [a,a], valid, [A,_,C], [a, terms(A, C)]).
conclusion(1, [a,a], invalid, [A,_,C], [n, terms(A, C)]).
conclusion(1, [i,a], valid, [A,_,C], [i, terms(A, C)]).
conclusion(1, [i,a], invalid, [A,_,C], [n, terms(A, C)]).
conclusion(1, [e,a], valid, [A,_,C], [o, terms(C, A)]).
conclusion(1, [e,a], invalid, [A,_,C], [e, terms(C, A)]).
conclusion(1, [o,a], valid, [A,_,C], [n, terms(A, C)]).
conclusion(1, [o,a], invalid, [A,_,C], [o, terms(A, C)]).

conclusion(1, [a,i], valid, [A,_,C], [n, terms(A, C)]).
conclusion(1, [a,i], invalid, [A,_,C], [i, terms(A, C)]).
conclusion(1, [i,i], valid, [A,_,C], [n, terms(A, C)]).
conclusion(1, [i,i], invalid, [A,_,C], [i, terms(A, C)]).
conclusion(1, [e,i], valid, [A,_,C], [o, terms(C, A)]).
conclusion(1, [e,i], invalid, [A,_,C], [e, terms(A, C)]).
conclusion(1, [o,i], valid, [A,_,C], [n, terms(A, C)]).
conclusion(1, [o,i], invalid, [A,_,C], [o, terms(A, C)]).

conclusion(1, [a,e], valid, [A,_,C], [e, terms(A, C)]).
conclusion(1, [a,e], invalid, [A,_,C], [n, terms(A, C)]).
conclusion(1, [i,e], valid, [A,_,C], [o, terms(A, C)]).
conclusion(1, [i,e], invalid, [A,_,C], [e, terms(A, C)]).
conclusion(1, [e,e], valid, [A,_,C], [n, terms(A, C)]).
conclusion(1, [e,e], invalid, [A,_,C], [e, terms(A, C)]).
```

```
conclusion(1, [o,e], valid, [A,_,C], [n, terms(A, C)]).
conclusion(1, [o,e], invalid, [A,_,C], [o, terms(A, C)]).

conclusion(1, [a,o], valid, [A,_,C], [n, terms(A, C)]).
conclusion(1, [a,o], invalid, [A,_,C], [o, terms(A, C)]).
conclusion(1, [i,o], valid, [A,_,C], [n, terms(A, C)]).
conclusion(1, [i,o], invalid, [A,_,C], [o, terms(A, C)]).
conclusion(1, [e,o], valid, [A,_,C], [n, terms(A, C)]).
conclusion(1, [e,o], invalid, [A,_,C], [o, terms(A, C)]).
conclusion(1, [o,o], valid, [A,_,C], [n, terms(A, C)]).
conclusion(1, [o,o], invalid, [A,_,C], [o, terms(A, C)]).

conclusion(2, [a,a], valid, [A,_,C], [a, terms(C, A)]).
conclusion(2, [a,a], invalid, [A,_,C], [n, terms(A, C)]).
conclusion(2, [i,a], valid, [A,_,C], [n, terms(A, C)]).
conclusion(2, [i,a], invalid, [A,_,C], [i, terms(A, C)]).
conclusion(2, [e,a], valid, [A,_,C], [e, terms(C, A)]).
conclusion(2, [e,a], invalid, [A,_,C], [n, terms(A, C)]).
conclusion(2, [o,a], valid, [A,_,C], [n, terms(A, C)]).
conclusion(2, [o,a], invalid, [A,_,C], [o, terms(C, A)]).

conclusion(2, [a,i], valid, [A,_,C], [i, terms(C, A)]).
conclusion(2, [a,i], invalid, [A,_,C], [n, terms(A, C)]).
conclusion(2, [i,i], valid, [A,_,C], [n, terms(A, C)]).
conclusion(2, [i,i], invalid, [A,_,C], [i, terms(C, A)]).
conclusion(2, [e,i], valid, [A,_,C], [o, terms(C, A)]).
conclusion(2, [e,i], invalid, [A,_,C], [e, terms(C, A)]).
conclusion(2, [o,i], valid, [A,_,C], [n, terms(A, C)]).
conclusion(2, [o,i], invalid, [A,_,C], [o, terms(C, A)]).

conclusion(2, [a,e], valid, [A,_,C], [o, terms(A, C)]).
conclusion(2, [a,e], invalid, [A,_,C], [e, terms(A, C)]).
conclusion(2, [i,e], valid, [A,_,C], [o, terms(A, C)]).
conclusion(2, [i,e], invalid, [A,_,C], [o, terms(C, A)]).
conclusion(2, [e,e], valid, [A,_,C], [n, terms(A, C)]).
```

```
conclusion(2, [e,e], invalid, [A,_,C], [e, terms(C, A)]).
conclusion(2, [o,e], valid, [A,_,C], [n, terms(A, C)]).
conclusion(2, [o,e], invalid, [A,_,C], [o, terms(C, A)]).

conclusion(2, [a,o], valid, [A,_,C], [n, terms(A, C)]).
conclusion(2, [a,o], invalid, [A,_,C], [o, terms(C, A)]).
conclusion(2, [i,o], valid, [A,_,C], [n, terms(A, C)]).
conclusion(2, [i,o], invalid, [A,_,C], [o, terms(C, A)]).
conclusion(2, [e,o], valid, [A,_,C], [n, terms(A, C)]).
conclusion(2, [e,o], invalid, [A,_,C], [o, terms(C, A)]).
conclusion(2, [o,o], valid, [A,_,C], [n, terms(A, C)]).
conclusion(2, [o,o], invalid, [A,_,C], [o, terms(A, C)]).

conclusion(3, [a,a], valid, [A,_,C], [n, terms(A, C)]).
conclusion(3, [a,a], invalid, [A,_,C], [a, terms(C, A)]).
conclusion(3, [i,a], valid, [A,_,C], [n, terms(A, C)]).
conclusion(3, [i,a], invalid, [A,_,C], [i, terms(A, C)]).
conclusion(3, [e,a], valid, [A,_,C], [e, terms(A, C)]).
conclusion(3, [e,a], invalid, [A,_,C], [n, terms(A, C)]).
conclusion(3, [o,a], valid, [A,_,C], [o, terms(A, C)]).
conclusion(3, [o,a], invalid, [A,_,C], [o, terms(C, A)]).

conclusion(3, [a,i], valid, [A,_,C], [n, terms(A, C)]).
conclusion(3, [a,i], invalid, [A,_,C], [i, terms(C, A)]).
conclusion(3, [i,i], valid, [A,_,C], [n, terms(A, C)]).
conclusion(3, [i,i], invalid, [A,_,C], [i, terms(A, C)]).
conclusion(3, [e,i], valid, [A,_,C], [o, terms(C, A)]).
conclusion(3, [e,i], invalid, [A,_,C], [e, terms(C, A)]).
conclusion(3, [o,i], valid, [A,_,C], [n, terms(A, C)]).
conclusion(3, [o,i], invalid, [A,_,C], [o, terms(A, C)]).

conclusion(3, [a,e], valid, [A,_,C], [e, terms(C, A)]).
conclusion(3, [a,e], invalid, [A,_,C], [n, terms(A, C)]).
conclusion(3, [i,e], valid, [A,_,C], [o, terms(A, C)]).
conclusion(3, [i,e], invalid, [A,_,C], [e, terms(C, A)]).
```

```
conclusion(3, [e,e], valid, [A,_,C], [n, terms(A, C)]).
conclusion(3, [e,e], invalid, [A,_,C], [e, terms(A, C)]).
conclusion(3, [o,e], valid, [A,_,C], [n, terms(A, C)]).
conclusion(3, [o,e], invalid, [A,_,C], [o, terms(C, A)]).

conclusion(3, [a,o], valid, [A,_,C], [o, terms(C, A)]).
conclusion(3, [a,o], invalid, [A,_,C], [o, terms(A, C)]).
conclusion(3, [i,o], valid, [A,_,C], [n, terms(A, C)]).
conclusion(3, [i,o], invalid, [A,_,C], [o, terms(C, A)]).
conclusion(3, [e,o], valid, [A,_,C], [n, terms(A, C)]).
conclusion(3, [e,o], invalid, [A,_,C], [o, terms(C, A)]).
conclusion(3, [o,o], valid, [A,_,C], [n, terms(A, C)]).
conclusion(3, [o,o], invalid, [A,_,C], [o, terms(A, C)]).

conclusion(4, [a,a], valid, [A,_,C], [i, terms(A, C)]).
conclusion(4, [a,a], invalid, [A,_,C], [a, terms(A, C)]).
conclusion(4, [i,a], valid, [A,_,C], [i, terms(A, C)]).
conclusion(4, [i,a], invalid, [A,_,C], [n, terms(A, C)]).
conclusion(4, [e,a], valid, [A,_,C], [o, terms(C, A)]).
conclusion(4, [e,a], invalid, [A,_,C], [e, terms(C, A)]).
conclusion(4, [o,a], valid, [A,_,C], [o, terms(C, A)]).
conclusion(4, [o,a], invalid, [A,_,C], [i, terms(A, C)]).

conclusion(4, [a,i], valid, [A,_,C], [i, terms(A, C)]).
conclusion(4, [a,i], invalid, [A,_,C], [n, terms(A, C)]).
conclusion(4, [i,i], valid, [A,_,C], [n, terms(A, C)]).
conclusion(4, [i,i], invalid, [A,_,C], [i, terms(A, C)]).
conclusion(4, [e,i], valid, [A,_,C], [o, terms(C, A)]).
conclusion(4, [e,i], invalid, [A,_,C], [o, terms(A, C)]).
conclusion(4, [o,i], valid, [A,_,C], [n, terms(A, C)]).
conclusion(4, [o,i], invalid, [A,_,C], [o, terms(C, A)]).

conclusion(4, [a,e], valid, [A,_,C], [o, terms(A, C)]).
conclusion(4, [a,e], invalid, [A,_,C], [e, terms(A, C)]).
conclusion(4, [i,e], valid, [A,_,C], [o, terms(A, C)]).
```

```
conclusion(4, [i,e], invalid, [A,_,C], [e, terms(A, C)]).
conclusion(4, [e,e], valid, [A,_,C], [n, terms(A, C)]).
conclusion(4, [e,e], invalid, [A,_,C], [e, terms(A, C)]).
conclusion(4, [o,e], valid, [A,_,C], [n, terms(A, C)]).
conclusion(4, [o,e], invalid, [A,_,C], [o, terms(C, A)]).

conclusion(4, [a,o], valid, [A,_,C], [o, terms(A, C)]).
conclusion(4, [a,o], invalid, [A,_,C], [o, terms(C, A)]).
conclusion(4, [i,o], valid, [A,_,C], [n, terms(A, C)]).
conclusion(4, [i,o], invalid, [A,_,C], [o, terms(A, C)]).
conclusion(4, [e,o], valid, [A,_,C], [n, terms(A, C)]).
conclusion(4, [e,o], invalid, [A,_,C], [o, terms(A, C)]).
conclusion(4, [o,o], valid, [A,_,C], [n, terms(A, C)]).
conclusion(4, [o,o], invalid, [A,_,C], [o, terms(C, A)]).
```

Bibliography

Baddeley, A. D. (1990). *Human memory: Theory and practice*. Lawrence Erlbaum Associates Ltd, Hove, UK.

Baker, R., Clark, R. A. J., and White, M. (2004). Synthesising contextually appropriate intonation in limited domains. In *Proceedings of the 5th ISCA Speech Synthesis Workshop*, pages 91–96, Pittsburgh, PA.

Baker, R. E. (2003). *Using Unit Selection to Synthesise Contextually Appropriate Intonation in Limited Domain Synthesis*. Master's thesis, University of Edinburgh.

Beckman, M. E. and Hirschberg, J. (1994). The ToBI annotation conventions. Available via the World Wide Web at http://www.ling.ohio-state.edu/~tobi/ame_tobi/annotation_conventions.html.

Black, A. and Taylor, P. (1997). Automatically clustering similar units for unit selection in speech synthesis. In *Eurospeech '97*, volume 2, pages 601–604.

Bolinger, D. (1965). *Forms of English*. Harvard University Press, Cambridge, Mass.

Büring, D. (1997). *The Meaning of Topic and Focus: The 59th Street Bridge Accent*. Routledge, London.

Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition*, **31**, 61–83.

Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics and point of view. In C. Li, editor, *Subject and topic*, pages 25–55. Academic Press, New York.

Chafe, W. L. (1987). Cognitive constraints on information flow. In R. S. Tomlin, editor, *Coherence and grounding in discourse*, pages 21–50. John Benjamins, Amsterdam.

Chapman, K. J. and Chapman, J. P. (1959). The atmosphere effect re-examined. *Journal of Experimental Psychology*, **58**, 220–226.

Chomsky, N. (1971). Deep structure, surface structure and semantic interpretation. In D. Steinberg and L. Jacobovits, editors, *Semantics*. Cambridge University Press, Cambridge.

Clark, R. A. J., Richmond, K., and King, S. (2004). Festival 2 - build your own general purpose unit selection speech synthesiser. In *Proceedings of the 5th ISCA Speech Synthesis Workshop*, pages 173–178, Pittsburgh, PA.

Dahl, O. (1974). Topic-comment structure revisited. In O. Dahl, editor, *Topic and Comment, contextual boundedness and focus*, volume 6 of *Papers in text linguistics*, pages 1–24. Helmut Buske, Hamburg.

de Carolis, B., Pelachaud, C., Poggi, I., and Steedman, M. (2004). APML, a mark-up language for believable behavior generation. In H. Prendinger, editor, *Life-like Characters. Tools, Affective Functions and Applications*, pages 65–85. Springer, Berlin.

Dickstein, L. S. (1975). Effects of instruction and premise order on errors in syllogistic reasoning. *Journal of Experimental Psychology: Human Learning and Memory*, **104**, 376–384.

Dickstein, L. S. (1981). The meaning of conversion in syllogistic reasoning. *Bulletin of the Psychonomic Society*, **18**, 135–138.

Downing, A. (1991). An alternative approach to theme: a systemic-functional perspective. *WORD*, **42**(2), 119–143.

Evans, J. S. B. T. (1989). *Bias in human reasoning: Causes and consequences*. Psychology Press, Hove, UK.

Evans, J. S. B. T., Barston, J. L., and Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory and Cognition*, **11**, 295–306.

Evans, J. S. B. T., Newstead, S. E., and Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Psychology Press, Hove, UK.

Evans, J. S. B. T., Handley, S. J., and Harper, C. (2001). Necessity, possibility, and belief: A study of syllogistic reasoning. *Quarterly Journal of Experimental Psychology*, **54A**, 935–958.

Eysenck, M. W. and Keane, M. T. (1995). *Cognitive psychology: A student's handbook*. Psychology Press, Hove, UK.

Firbas, J. (1992). *Functional sentence perspective in written and spoken communication*. Cambridge University Press.

Ford, M. (1995). Two modes of mental representation and problem solution in syllogistic reasoning. *Cognition*, **54**, 1–71.

Foster, M. E. (2004). Generating output in the comic multimodal dialogue system. Talk presented at W3C Workshop on Multimodal Interaction, Sophia Antipolis, France.

Fries, C. (1983). On the status of theme in english. In J. Petöfi and E. Sözer, editors, *Micro- and macro-connexity of texts*, pages 116–152. Buske Verlag, Hamburg.

Gilhooly, K. J., Logie, R. H., Wetherick, N. E., and Wynn, V. (1993). Working memory and strategies in syllogistic-reasoning tasks. *Memory and Cognition*, **21**, 115–124.

Grice, H. P. (1975). Logic and conversation. In P. Cole and J. L. Morgan, editors, *Speech Acts*, volume 3 of *Syntax and Semantics*, pages 41–58. Academic Press, New York.

Grice, H. P. (1989). *Studies in the way of words*. Harvard University Press, Cambridge, Mass.

Grosz, B. J., Joshi, A. K., and Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, **2**, 203–225.

Gundel, J. (1974). *The role of topic and comment in linguistic theory*. Ph.D. thesis, University of Texas.

Gundel, J. (1980). Zero np-anaphora in russian: a case of topic-prominence. In *Chicago Linguistics Society 16: Papers from the Parasession on Anaphora*, pages 139–146.

Gundel, J. (1988). Universals of topic-comment structure. In M. Hammond, E. Moravcsik, and J. Wirth, editors, *Studies in syntactic typology*, pages 209–239. John Benjamins, Amsterdam.

Gundel, J. (1999). Topic, focus and the grammar-pragmatics interface. In *Proceedings of the 23rd Annual Penn Linguistics Colloquium*, volume 6.1 of *U. Penn Working Papers in Linguistics*.

Gussenhoven, C. (1983). Focus, mode and the nucleus. *Journal of Linguistics*, **19**, 377–417.

Halliday, M. A. K. (1967). Notes on transitivity and theme in english, part ii. *Journal of Linguistics*, **3**, 199–244.

Heim, I. (1982). *The semantics of definite and indefinite noun phrases*. Ph.D. thesis, University of Massachusetts.

Henle, M. (1962). On the relations between logic and thinking. *Psychological Research*, **69**, 366–378.

Hockett, C. F. (1958). *A Course in Modern Linguistics*. Macmillan, New York.

Inder, R. (1987). *the computer simulation of syllogism solving using restricted mental models*. Ph.D. thesis, University of Edinburgh.

Jackendoff, R. (1972). *Semantic interpretation in generative grammar*. MIT Press, Cambridge, Mass.

Janis, I. L. and Frick, F. (1943). The relationship between attitudes toward conclusions and errors in judging logical validity of syllogisms. *Journal of Experimental Psychology*, **33**, 73–77.

Johnson-Laird, P. and Byrne, R. M. J. (1991). *Deduction*. Lawrence Erlbaum, Hillsdale, NJ.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference and consciousness*. Harvard University Press, Cambridge, MA.

Johnson-Laird, P. N. and Bara, B. G. (1984). Syllogistic inference. *Cognition*, **16**, 1–61.

Johnson-Laird, P. N. and Steedman, M. (1978). The psychology of syllogisms. *Cognitive Psychology*, **10**, 64–99.

Karttunen, L. (1976). Discourse referents. In J. McCawley, editor, *Notes from the Linguistic Underground*, volume 7 of *Syntax and Semantics*, pages 363–385. Academic Press, New York.

Karttunen, L. and Peters, S. (1979). Conventional implicature. In C.-K. Oh and D. Dineen, editors, *Presupposition*, volume 11 of *Syntax and Semantics*, pages 1–56. Academic Press, New York.

Klauer, K. C., Musch, J., and Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review*, **107**, 852–884.

Kruijff-Korbayová, I. and Steedman, M. (2003). Discourse and information structure. *Journal of Logic, Language and Information*, **12**, 249–259.

Ladd, D. R. (1996). *Intonational Phonology*. Number 79 in Cambridge Studies in Linguistics. Cambridge University Press.

Lambrecht, K. (1994). *Information structure and sentence form: topic, focus and the mental representation of discourse referents*. Cambridge University Press, Cambridge.

Monaghan, P. and Stenning, K. (1998). Effects of representational modality and thinking style on learning to solve reasoning problems. In *Proceedings of the 20th annual meeting of the Cognitive Science Society*, pages 716–721, Madison, Wisconsin.

Morgan, J. J. B. and Morton, J. T. (1944). The distortion of syllogistic reasoning produced by personal convictions. *Journal of Psychology*, **20**, 39–59.

Newstead, S. E. (1989). Interpretational errors in syllogistic reasoning. *Journal of Memory and Language*, **28**, 78–91.

Newstead, S. E. (1995). Gricean implicatures in syllogistic reasoning. *Journal of Memory and Language*, **34**, 644–664.

Newstead, S. E. and Griggs, R. A. (1999). Premise misinterpretation and syllogistic reasoning. *Quarterly Journal of Experimental Psychology*, **52A**, 1057–1075.

Oakhill, J. V. and Johnson-Laird, P. N. (1985). The effects of belief on the spontaneous production of syllogistic conclusions. *Quarterly Journal of Experimental Psychology*, **37A**, 553–569.

Oakhill, J. V., Johnson-Laird, P. N., and Garnham, A. (1989). Believability and syllogistic reasoning. *Cognition*, **31**, 117–140.

Pierrehumbert, J. and Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, and M. Pollack, editors, *Intentions in Communication*, pages 271–312. MIT Press, Cambridge, MA.

Politzer, G. (2004). Some precursors of current theories of syllogistic reasoning. In K. Manktelow and M. C. Chung, editors, *Psychology of Reasoning. Theoretical and historical perspectives*, pages 213–240. Psychology Press, Hove and New York.

Polk, T. A. and Newell, A. (1995). Deduction as verbal reasoning. *Psychological Review*, **102**(3), 533–566.

Prince, E. F. (1986). On the syntactic marking of presupposed open propositions. In *Chicago Linguistics Society 22: Parasession papers*, pages 208–222.

Quayle, J. D. and Ball, L. J. (2000). Working memory, metacognitive uncertainty, and belief bias in syllogistic reasoning. *Quarterly Journal of Experimental Psychology*, **53A**(4), 1202–1223.

Roberts, M. J., Newstead, S. E., and Griggs, R. A. (2001). Quantifier interpretation and syllogistic reasoning. *Thinking and Reasoning*, **7**(2), 173–204.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, **1**, 75–116.

Schmerling, S. F. (1976). *Aspects of English sentence stress*. University of Texas Press, Austin.

Selkirk, E. (1984). *Phonology and Syntax*. MIT Press, Cambridge, MA.

Sells, S. B. (1936). The atmosphere effect: An experimental study of reasoning. *Archives of Psychology*, (200), 1–72.

Sgall, P. (1967). Functional sentence perspective in a generative description. *Prague Studies in Mathematical Linguistics*, **2**, 203–225.

Sgall, P., Hajičová, E., and Panevová, J. (1986). *The Meaning of the Sentence in its Semantic and Pragmatic Aspects*. Reidel, Dordrecht.

Shore, J. (1988). Interactive signal processing with UNIX. *Speech Technology*, **3**. (Mar/Apr).

Steedman, M. (1991). Structure and intonation. *Language*, **67**, 260–296.

Steedman, M. (2000). Information structure and the syntax-phonology interface. *Linguistic Inquiry*, **34**, 649–689.

Steedman, M. (2004). Using APML to specify intonation. Unpublished tutorial paper.

Stenning, K. (1996). Embedding logic in communication: lessons from the logic classroom. In J. van Benthem, F. H. van Eemeren, R. Grootendorst, and F. Veltman, editors, *Logic and Argumentation*, pages 227–240, Amsterdam. Royal Netherlands Academy of Arts and Sciences.

Stenning, K. and Cox, R. (1995). Attitudes to logical independence: traits in quantifier interpretation. In *Proceedings of the 17th Annual Conference of the Cognitive Science Society*. Lawrence Erlbaum Associates, Pittsburgh.

Stenning, K. and Cox, R. (2006). Reconnecting interpretation to reasoning through individual differences. *Quarterly Journal of Experimental Psychology*.

Stenning, K. and van Lambalgen, M. (2004). A little logic goes a long way: basing experiment on semantic theory in the cognitive science of conditional reasoning. *Cognitive Science*, **28**(4), 481–530.

Stenning, K. and van Lambalgen, M. (2005). A working memory model of relations between interpretation. *Cognitive Science*.

Stenning, K. and Yule, P. (1997). Image and language in human reasoning: A syllogistic illustration. *Cognitive Psychology*, **34**, 109–159.

Stenning, K., Yule, P., and Cox, R. (1996). Quantifier interpretation and syllogistic reasoning. In *Proceedings of the 18th Annual Conference of the Cognitive Science Society*, pages 201–206. Lawrence Erlbaum Associates, Pittsburgh.

Störring, G. (1908). Experimentelle untersuchungen zu einfachen schlußprozessen. *Archiv für die Gesamte Psychologie*, (11), 1–127.

Talkin, D. (1989). Looking at speech. *Speech Technology*, **4**. (Apr/May).

Taylor, P., Black, A., and Caley, R. (1998). The architecture of the festival speech synthesis system. In *Proceedings of the Third ESCA Workshop in Speech Synthesis*, pages 147–151.

Vallduví, E. (1992). *The informational component*. Outstanding dissertations in linguistics. Garland, New York.

Vallduví, E. and Vilkuna, M. (1998). On rheme and kontrast. In P. Culicover and L. McNally, editors, *The Limits of Syntax*, volume 29 of *Syntax and Semantics*, pages 79–108. Academic Press, San Diego.

van Deemter, K. (1992). What's new? semantic notions of 'new information' for intonational focusing. IPO Eindhoven manuscript.

van Oosten, J. (1986). *The nature of subjects, topics, and agents*. Indiana University Linguistics Club, Bloomington.

Walker, M., Joshi, A. K., and Prince, E. F., editors (1998). *Centering Theory in Discourse*. Oxford University Press, Oxford.

Ward, G. and Hirschberg, J. (1985). Implicating uncertainty: the pragmatics of fall-rise intonation. *Language*, **61**, 747–776.

Wason, P. and Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. Batsford, London.

Wason, P. C. (1966). Reasoning. In B. M. Foss, editor, *New Horizons in Psychology*, volume I. Penguin.

Wilkins, M. C. (1928). The effect of changed material on ability to do formal syllogistic reasoning. *Archives of Psychology*, (102).

Woodworth, R. S. and Sells, S. B. (1935). An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology*, **18**, 451–460.